An Explanation of the Determinants of Prescribing Expenditure in the General Medical Services Scheme in Ireland with an Application to Budget Setting

Brendan McElroy

PhD Dissertation

The University of Edinburgh

2003

Dedicated to my parents, Mary and Michael, for their unstinting love, support and inspiration

CONTENTS

LIST OF TABLES

LIST OF FIGURES

DECLARATION

I hereby and formally declare that this work, entitled *An explanation of the determinants of prescribing expenditure in the General Medical Services scheme in Ireland with an application to budget setting*, has been composed by myself; that it has not been accepted in any previous application for a degree and that the work of which is it a record has been done by myself.

Signed

Brendan McElroy

January 2003

DECLARATION

I hereby and formally declare that this work, entitled *An explanation of the determinants of prescribing expenditure in the General Medical Services scheme in Ireland with an application to budget setting*, has been composed by myself; that it has not been accepted in any previous application for a degree and that the work of which is it a record has been done by myself.

Signed

Brendan McElroy
January 2003

ACKNOWLEDGEMENTS

I wish to acknowledge the support of everyone who helped me in the completion of this thesis. In particular I would like to thank the following:

- Dr. John Forbes for his expert advice, encouragement, interest and constant availability
- The Health Research Board and Voluntary Health Insurance for providing generous financial support for this project
- Members of the Department of Economics, University College Cork for making this experience both productive and enjoyable. Special mention to the members of the PhD Workshop for their support and advice, Professor Connell Fanning for his encouragement and enthusiasm, and Mary Maguire and Dympna O'Donovan for their administrative support
- Professor Colin Bradley for many interesting discussions and insightful comments
- Carolyn Macdougall for her administrative support
- For their specialist academic advice, Matt Sutton, Peter Lock, Partha Deb, Willard Manning, Gary King, John Carlin, Mark Hornbrook and his team at Kaiser-Permanente, Kirsten Major and Leida Lamers
- Delegates at the many conferences and workshops at which parts of this research were presented. Special thanks to the delegates at the 11th European Workshop on Health Economics and Econometrics (2001) and the Health Economists' Study Group meeting (September, 2001)
- Members of the GMS (Payments) Board, Southern Health Board, South Eastern Health Board and North Eastern Health Board who provided data and responded to my many queries. Special thanks to Ciaran Coughlan, Paddy Burke, Louise Hannon, Vera Matthews, Des Cassidy, Cait Keyes and Nuala Kelleher
- GPs 'John' and 'Joe' for agreeing to be case study material
- My family and friends for making it such good fun. Special thanks to Ann and Helena, Mark and Cian, Claire and Jeremy, and Denise.

ABSTRACT

The Indicative Drug Target Scheme was introduced to the General Medical Services (GMS) scheme in 1993 with a view to encourage more rational and efficient prescribing. Each GP practice's drug budget is determined chiefly by the number of people on the GP's GMS panel adjusted for national age-related average prescribing expenditure. This thesis examines the factors associated with variations in GMS prescribing expenditure, proposes a number of alternative ways of constructing GMS drug budgets and examines the potential budgetary consequences.

A unique dataset of individual-level and GP-level factors is constructed, including the first research application of an administrative database of demographic, socio-economic and access-related variables, the generation of chronic illness indicators, and the application of new measures of GP prescribing style. Multiple imputation of missing values and imputation of income are two additional innovations. Drawing on recent advances in risk-adjustment and the microeconometrics of health care utilisation, various specifications of an expenditure function are examined, given the skewed distribution of prescribing expenditure. These include logarithmic transformations, generalised linear models and finite mixture models. Quantile regression and outlier identification techniques are used for exploratory data analysis and to assist model specification.

The principal determinants of prescribing expenditure are chronic illness, disability and age. Access to services and GP characteristics also have important effects. A number of competing models of budget setting are tested for predictive performance and distributive consequences. Most alternatives are an improvement on the current model. Compared to the current approach, the preferred model has greater explained variance, lower prediction error, is more pro-poor and has lower prediction error for vulnerable groups such as the disabled and the chronically ill.

1. INTRODUCTION

1.1 INTRODUCTION AND STRUCTURE

The purpose of this study is to analyse the determinants of prescribing expenditure by recipients of General Medical Services (GMS) in Ireland. Having quantified the effect of each determinant, we then estimate each individual's need for prescription medicine. In doing so, we wish to inform the setting of GMS prescribing budgets.

General Practitioners (GPs) in Ireland with GMS contracts receive an indicative budget for their GMS prescribing. A fraction of budgetary savings can be retained for approved practice developments. Currently, GMS prescribing budgets are set on the basis of an age adjustment to each General Practitioner's (GP's) patient list. However, previous research has indicated a number of other factors that affect health care utilisation, including gender, socio-economic status, health status, access to health services and characteristics of the health care providers. This study examines the effects of these factors as well as age on prescribing expenditure. Some of these factors are indicators of prescribing need and can be included in the GMS budget-setting formula. Others are measures of differential access to services or differential treatment of individuals by health care providers and need to be controlled for when estimating the independent effect of needs variables. As such, we wish to provide a formula for budget setting that is more responsive to patient need, has better predictive ability and is more equitable than the current one.

This study falls into the broad category of health economic research known variously as risk-adjustment or weighted capitation modelling (words that we use interchangeably), which examines the characteristics of a population of health service users, be they residents of a Health Board area or subscribers to a health plan, and assigns weights to each relevant characteristic that reflect the effect of each characteristic on the need for health care.

In the remainder of this chapter we describe the current GMS regime in Ireland and the budget setting mechanism in more detail, in order to place the current study in its policy context. We also outline the two principal approaches to weighted capitation. The 'normative' approach seeks to measure the burden of illness in the population

and determine the need for health care on this basis. The 'empirical' approach infers the need for health care from the relationship between health service utilisation and such covariates as socioeconomic and demographic characteristics of the population and the supply of health services. This study focuses on the latter.

Chapter 2 examines the empirical approach to risk-adjustment in more detail. We begin by reviewing theoretical models of utilisation of health care. First, there are health care utilisation models that focus on the choices of the individual, such as Grossman's (1972) demand for health care model. Second, there are models that focus on the choices of the physician. We then review empirical models of utilisation. Drawing on both theoretical and empirical research we generate expected effects for the set of variables included in the study. Next, an overview of the principal econometric issues is provided. These refer to awkward features of health care datasets such as truncation at zero expenditure, a high proportion of the sample having zero utilisation (representing non-users) and a positive skew. We critically review the large number of remedies that have been proposed, including the popular two-part models, which separately generate estimates of one's probability of any use of health care and the extent of that use. Having considered the issues involved in generating a utilisation function that is theoretically and statistically acceptable, we then address additional considerations in constructing GMS prescribing budgets. These include the relationship between utilisation and need; the distributional consequences of applying a new budget setting formula and the effects of risk exposure faced by the budget holder under different budget setting regimes.

Chapter 3 describes the dataset, assesses its quality and describes imputation methods. A unique dataset has been constructed for this study, containing data on over 400,000 users of GMS services. These data include their prescribing expenditure over a 12-month period, their age and gender as well as many indicators of their socio-economic circumstances, health status, access to health services and characteristics of their GPs. The dataset has been generated partly from the Health Board's Medical Card Registers, which is the first time this source has been used for health services research as far as we are aware. In addition, prescribing data from the GMS (Payments) Board has been used to generate individual-level indicators of chronic illness, GP-level indicators of prescribing style as well as each individual's

prescribing expenditure. Finally, the GMS (Payments) Board provided data on characteristics of the GPs studied and their practices. A number of checks of the reliability of the data are reported. Summary statistics from the variables used in this study are compared with those for similar variables in other studies to assess the external validity of the dataset. Measurement error in chronic illness indicators is rigorously examined. The problem of missing values, especially in the disability field, is addressed and a best-practice solution is applied. As part of model evaluation, we wish to assess each model's distribution of prescribing resources by income. We estimate a model of income using the Household Budget Survey 1999/2000 (2002) and use this model to predict income for every individual in the study. Finally, we describe a set of variables on vulnerable groups that are not universally recorded, so they cannot be used in model estimation, but that are used for model evaluation.

There is no consensus in previous econometric literature on a number of model specification issues. Some of these are addressed in Chapter 4. First, previous research is unclear about the correct specification of the relationship between age and health care utilisation. We examine the effects of a number of specifications on model performance. Second, the distribution of the response variable is described in detail such that the extent of the spike at zero expenditure and the positive skew are better understood. The next three sections then describe aspects of this distribution. The effect of outliers on model consistency is assessed using outlier identification techniques and the application of outlier-robust regression techniques. The effect of the determinants of prescribing expenditure at different points on the conditional distribution is then examined using quantile regression. Next, as an alternative to one-part or two-part models of health care utilisation, finite mixture models are explored. These can be used to test the hypothesis that the response variable consists of a mixture of distributions, rather than just one. On the basis of the analysis in these three sections, we conclude that the removal of high cost outliers from the dataset is warranted but that the application of outlier-robust regression or finite mixture models is not.

Chapter 5 presents and discusses the results of a number of competing empirical estimators of health care utilisation. These include four one-part estimators, which

use ordinary least squares (OLS), and two two-part estimators. The one-part models include the current budget-setting model and three models with differential restrictions on variable inclusion. The two-part models differ in their specification of the second part. The first two-part model uses a log-transformed response variable for the second part. The second one uses a generalised linear model for the second part. The results are compared across models and are critically discussed in the context of previous research.

The aim of the study is not just to explain the process that generates any particular level of prescribing expenditure (which is reported in Chapter 5), but also to consistently and precisely predict prescribing expenditure, in order that budgets can best reflect the relative needs of the GMS population. In addition, we wish budget setting models not only to reflect relative needs, but to distribute resources equitably. Thus, Chapter 6 examines the predictive ability of each competing model, as well as their distributive consequences. While many risk-adjustment studies examine average prediction error only, we also test for prediction bias and examine the distribution of prediction error with respect to certain groups, including the poor, the chronically ill, people living alone, asylum seekers and early school leavers.

Chapter 7 aggregates the results up to GP-level and assesses the predictive and distributive ability of each model after aggregation. In addition, the risk exposure of the GP as budget holder is described at various sizes of GP practice and with various procedures for treating outliers. The consequences of each model for the distribution of GMS prescribing budgets is then described. Finally, two case studies of the effect of current GMS indicative drug budgets and the changes that would occur with the application of the findings of this study for two GP budget holders are described. Chapter 8 offers conclusions, recommendations and suggests areas for further research.

1.2 INDICATIVE DRUG TARGET SCHEME

The GMS scheme was set up under the Health Act 1970. GMS eligibility (also known as medical card entitlement) is granted to *'persons who are unable without undue hardship to arrange general practitioner services for themselves and their*

dependants' (CSO, 1999:100). The income thresholds for various groups are outlined in Table 1.1:

Table 1.1

Income Criteria for Eligibility for GMS for selected groups, March 2000

	Weekly Income Threshold (IR£)			
Group	Age < 66	Age 66-69	Age 70-79	Age 80+
Single person living alone	93.50	101.50	168.50	177.50
Single person living with family	83.00	88.00	146.00	152.00
Married Couple	135.00	151.00	252.50	265.00

Source: Southern Health Board, 2000

The majority of medical cards are granted on the basis of income and age. There are also discretionary medical cards, granted to hardship cases. Since July 2001, all those aged over 70 are entitled to a medical card. Our study period is up to May 2001, so the analysis is unaffected by this change in policy.

Whilst income thresholds for GMS eligibility are clearly quite low and therefore GMS patients can be considered to be in a broadly similar socio-economic group, all GMS recipients are not, in fact, in the lowest income categories due to the granting of discretionary medical cards. Nolan and Russell (2001), in an analysis of the distribution of medical card benefits by income decile, found that 60.5% of benefits went to the lowest three income deciles. The remaining 39.5% (amounting to 467,503 people in December 1998) were from higher income deciles, including 14.4% who were from the top half of the income distribution. Therefore there are likely to be variations in socio-economic characteristics within the GMS patient population with consequent variations in the need for prescription medicine.

Up until 1993 GMS prescribing was financed through retrospective reimbursement. In 1993 an Indicative Drug Target Scheme (IDTS) was introduced. The scheme provided for *"the calculation of monetary prescribing targets for each General Practitioner (GP) taking into consideration the make-up of his/her patient panel"* (Dept of Health and Children, 1997:iii), with a view to encouraging more rational and economic prescribing.

The objective of the IDTS is to give GPs an incentive to economise on their prescription of medicine. GPs who make savings on their drug budget are allowed to

retain a certain percentage of the savings to spend on approved practice developments. The manner in which budgets have been calculated and the exact percentage of savings that the GP are allowed to retain has varied over time.

In 1993 national age-related average prescribing expenditures (the NARA) for each of seven age categories in 1992 were calculated and applied to each GP's practice list. Their target that year was either their 1992 expenditure or the NARA applied to their panels, whichever was the lower, plus an inflation uplift of 7.5%, plus an allowance for patients on high expenditure drugs. Aside from some minor adjustments, the 1994 targets were the 1993 targets inflated by 6.4%. In 1995 GPs were split into three categories, A to C, where the A category referred to GPs whose 1994 expenditure was more than 105% of the 1994 NARA applied to their panel, C referred to GPs whose 1994 expenditure was less than 95% of the 1994 NARA applied to their panel and B referred to the group in between. GPs in the A category essentially got their 1994 expenditure as their 1995 target while B and C category GPs essentially got their 1994 expenditure inflated by 5% as their 1995 target. In 1996 categories A to C were further split into A1 to A3, B1 and B2 and C1 to C3, based on their deficits or surpluses and their expenditure relative to the 1995 NARA applied to their panels. These categories have been retained since 1996. The 1996 target for A1 GPs was essentially their 1995 target, while the target for A2 to C3 GPs was essentially their 1995 target plus a flat increase of IR£8 per patient. The 1997 target for A1 GPs was essentially their 1996 target, while other categories of GP essentially received their 1996 target plus a flat increase of IR£12 per patient. The 1998 target for A1 GPs was essentially their 1997 target or the 1997 NARA applied to their panel, whichever was the greater. Other categories of GP essentially received their 1997 expenditure plus an age-adjusted increase per patient ranging from IR£3 for 5-15 year olds to IR£25 for over 65s. The use of the word 'essentially' is because additional considerations of lesser consequence prevailed as well. For instance, certain high cost drugs were exempt, as were so-called 'budget neutral' drugs, which are drugs that are either high cost or whose prescription is promoted as part of health policy, such as lipid-lowering drugs as part of the 2000 cardiovascular strategy. Further details of the budget setting methodology are available in Joyce-Cooney (1999: Appendix A).

The budget setting methodology appears to be based on the NARA subject to a number of ad hoc adjustments, which have been motivated chiefly by the extent of the deviation of some GPs expenditure from the NARA[1]. Since 1998 similar ad hoc adjustments have been applied. These adjustments have diminished the transparency of the budget setting methodology, such that in 2001 the Irish Medical Organisation, with funding from the GMS (Payments) Board, established a review group to outline the history of the budget setting methodology and suggest ways of improving it.

There have also been adjustments to the extent to which individual GPs could retain savings for practice development. In 1993 and 1994 50% of savings could be retained by the practice, and 50% was spent by the Health Board for the overall development of general practice. Since then, the proportion of savings that could be retained has varied from 40% to 100% depending on the category of GP and year in question (Joyce-Cooney, 1999).

Analysis of the IDTS has focused on the extent to which it has affected efficiency in the prescription of medicine (Department of Health and Children, 1997). A commissioned review of the scheme in 1997 focused on *"quality of patient care in the GMS with particular reference to changes in prescribing patterns which have occurred since the introduction of the scheme"* (Murphy, 1997). The review examined the degree to which prescribing was 'rational', 'appropriate' or 'optimal'. It found that there was a greater level of generic prescribing since the introduction of the scheme, with no discernible negative effects on quality of prescribing. On the other hand, the Comptroller and Auditor General (Department of Health and Children, 1997) found that those GPs whose expenditure was below target in the four years to 1996 made savings of IR£18.3million, while those GPs whose expenditure exceeded their target had deficits of IR£43 million. Over that period only 5% of GPs continually made savings, while 27% of GPs continually had deficits. However the percentage of GPs who either saved or whose overspend was less than 10% of their target grew from 64.7 to 73.2% over the four year period. Therefore, if we use expenditure below- or on-target as a measure of the scheme's success, improvements have been modest.

[1] Personal communication with member of Irish Medical Organisation Executive, 2002.

By comparison with work on efficiency of the scheme, the extent to which budget allocations accurately reflect the needs of patients, and therefore the extent to which it promotes equity in the distribution of prescribing resources has been underexamined. The criteria underpinning the budget setting process appear to be a combination of needs (the NARA is an attempt is weight panel size for age-related differences in prescribing need) and risk-sharing (prescribing bill for budget-neutral patients and patients on high-cost medicine is paid by the GMS (Payments) Board, who, because of its greater population, has a more predictable level of expenditure on this type of patient than an individual general practice), along with a number of ad hoc adjustments (differential treatment of uplifts for inflation for different categories of GP). This study focuses primarily on the first consideration – incorporating need in the budget setting process. However, we also consider the role of risk-sharing and the potential for cost shifting or risk selection in determining the budget setting formula.

1.3 APPROACHES TO RISK-ADJUSTMENT

There are two broad approaches to risk-adjustment, 'normative' and 'empirical'. These labels are somewhat unsatisfactory, implying that the 'normative' approach is concerned only with value judgements, and the 'empirical' approach contains none. However, they are commonly used in the risk-adjustment literature and we stick to them here. The normative approach is reviewed first. The empirical approach is then outlined and an explanation of why it is adopted in this study is furnished.

1.3.1 Normative

The normative approach, sometimes known as the epidemiological approach, to weighted capitation modelling involves collecting data on ill-health in the population and determining the resource requirements in providing health care to deal with the identified ill-health. Most applications that include the normative approach are a hybrid of the normative and empirical approaches, where age- or age- and gender-related utilisation rates are adjusted by a needs score measured using data on ill-health. Potential applications of the normative approach can be categorised into those based on data on reported illness, those based on standardised mortality ratios (SMRs) and those based on Disability Adjusted Life Years (DALYs). Two of the principal criticisms of the normative approach is the arbitrary application of resource

weights to levels of need and second, the focus on burden of illness as a measure of need, rather than capacity to benefit (Williams, 1998). One remedy to the first of these is also considered here.

1.3.1.1 Reported Illness

Estimates of the prevalence of illness can be calculated using data on reported illness, such as through questionnaires administered by public-sponsored statistics offices such as the Census of Population and the Quarterly National Household Survey, or based on records of absence from work due to illness.

Owing to known under-utilisation of primary care services by deprived sub-groups of the population, the Swedish primary care model did not employ data on utilisation of services to determine need (Rice and Smith, 1999:75). Rather need was measured as the proportion of the population in each small area aged 16 to 64 suffering permanent sickness or with at least 30 days sickness in a one year period. This measure was regressed on three explanatory variables: proportion with low and middle income (<200,000 Swedish Kroner); proportion aged 45-64 married or divorced and proportion born abroad or foreign citizens. Predicted need scores were thus determined, which were used to adjust age weights to calculate a need score for each small area, making it a hybrid of normative and empirical approaches. This need score was used as the basis for resource allocation, with an assumed 1:1 relationship between needs score and need for health care resources. Le Grande (1978) used a question on 'limiting long-standing illness' from the General Household Survey for England as a measure of need in an analysis of the geographic distribution of national health service (NHS) resources in England, with a need of one for those reported having a limiting long-standing illness and a need of zero for those who did not.

There are a number of drawbacks with this approach. First, the assumption of a 1:1 relationship between need and need for health care resources, implicit in both the above studies has little basis in evidence. Indeed the choice of resource weights attached to various states of ill-health can have significant effects on the final distribution of health care resources. For instance Sutton et al. (1999a) categorised the population of Greater Glasgow Health Board on the basis of reported ill-health.

There was a substantial difference in resources allocated to different areas when a weight of two was attached to the 'sick' and one to the 'healthy' compared to weights of one to the 'sick' and zero to the healthy. This is probably the chief drawback of the normative approach.

Four further drawbacks apply to the use of limiting long-standing illness. First is that the dichotomous nature of the variable may mean that it poorly captures actual health status. Second, since limiting long-term illness is self-reported, biases may arise. It has been shown to better reflect physical illness rather than mental or social well-being, while individual perceptions of what constitutes long-term illness can vary with age, gender and recent visits to a GP (Sutton et al., 1999b). Third, if it became the basis for resource allocation, there would be an incentive on a self-reported questionnaire to report the presence of long-term illness. Fourth, as a variable recorded in national household surveys or the Census of Population, it may lack recency.

1.3.1.2 Standardised Mortality Ratio

The SMR standardises the age- and gender-specific death rates in an area against the expected death rates, given the area's age profile. SMRs of greater than 100 represent areas with greater than expected death rates, given their age and gender profiles. The original weighted capitation formula used in England, known as RAWP (Resource Allocation Working Party) and its Scottish counterpart SHARE (Scottish Health Area Resource Equalisation) adopted a hybrid of the normative and empirical approaches similar to the Swedish primary care model described above. The age- and gender-specific utilisation weights were generated for each area and these were weighted for 'need' using the SMR for those aged 0-64. Again, the resource weight given to SMR in SHARE and RAWP was 1:1, meaning that an area with an SMR of 120 would get 20% more resources, all other things being equal. Meanwhile, the Scottish Executive (1999:48) found that the bivariate relationship between use of acute services and SMR for people aged 0-64 in Greater Glasgow Health Board was:

$$\text{Use} = 68 + 0.32 * \text{SMR64},$$

where SMR64 is the SMR for those aged 0-64. Again the lack of evidence underpinning the assumption of a 1:1 relationship between need, as measured by the

SMR, and need for health care resources is the principal drawback to the SHARE and RAWP approaches.

Two additional problems arise with the use of SMR as a measure of need. First, it measures mortality not morbidity. Mortality in an area is often a good indicator of certain types of morbidity in the area, especially those types of morbidity that can kill such as coronary heart disease and cancer, it is not as good an indicator of other types of morbidity or need for health care resources in areas such as maternity services. Second, it is only available at county level in all Irish counties except Dublin, where it is available at postal district level. This is too great a level of aggregation to detect differences in need at GP practice level.

1.3.1.3 Disability Adjusted Life Years

A further development on the normative approach, which has been underexplored by researchers, involves calculating the burden of illness in each area using Disability-Adjusted-Life Years (DALYs). Drawing on the pioneering work of Murray and Lopez (1996), the DALYs approach breaks down the ill-health related to each condition into years of life lost and years of life in disability. Years of life lost are calculated using condition-specific life tables. The degree of disability associated with each condition is based on weightings given by expert panels. Future levels of disability can be discounted to present value terms and an age weighting can be applied. By combining years of life lost with years of life spent in disability (multiplied by the disability weight), one can produce an estimate of the burden of illness.

Whilst offering a more sophisticated measure of illness than limiting long-term illness, for instance, it nevertheless offers no solutions to the weights to attach to each level of ill-health in order to allocate resources. As such, it is likely to suffer from the same charge that the arbitrary choice of weight will have a large effect on final allocations. Other criticisms also apply. First, the choice of disability weights is based on expert opinion, rather than a more robust evidence base. Second, given the number of conditions to consider, aggregation is necessary. Therefore, the sensitivity of the disability weights attached to each condition is open to question. For instance, mild dementia and moderate dementia get disability weights of 0.6,

while severe dementia gets a weight of 0.85 (Murray and Lopez, 1996). In reality, moderate dementia is more 'disabling' than mild dementia and at any rate, dementia is experienced on a continuum rather than in discrete steps.

Third, the prevalence of the multitude of conditions considered imposes significant data requirements. Epidemiological data that are robust, comprehensive, timely and locally sensitive are rarely available. Even where these data are reasonably comprehensive, such as with the National Cancer Registry in Ireland, the data do not provide the necessary specificity in measuring health status. For instance, although the stage of the cancer tumour has significant implications for health care need, tumour staging is not recorded for many cancers, leaving us short on evidence of relative severity and relative health care need. The epidemiological data that are available have been estimated with varying degrees of sophistication, depending on available research resources. In many instances prevalence estimates are not applicable, even at health board level.

Fourth, the whole approach is based on a medical model of health, with the potential undervaluation of social well-being. Finally, the entire DALYs approach has been subject to numerous criticisms, most notably that it is based on the 'wrong' measure of need for resource allocation purposes, that is it uses the burden of illness definition not the capacity to benefit definition. In order to maximise the benefit of scarce health care resources, they should be deployed on conditions where treatment is most effective, and not where the burden of illness is greatest. To deploy resources on the treatment of an incurable illness that happens to be very prevalent is inefficient, given that those resources could be used effectively in treating another illness (Williams, 1998).

1.3.1.4 Guidelines and Resource Weights

In response to the criticism of the normative approach regarding the arbitrary use of resource weights, it is argued that for any given condition at any given stage, it should be possible to develop a set of best practice guidelines for its treatment. The cost of implementing such guidelines would then represent the resource weight attached to the prevalence of the condition in a region for resource allocation purposes.

The first difficulty with such an approach is affordability. Given limited resources and seemingly infinite demand for health care, it is not possible to adopt best practice guidelines for all conditions. Perhaps, therefore, the best practice approach should only be adopted to spend growth monies, with the rest of the budget being allocated on the basis of, say, an empirical model. Again, difficulties arise. Decision-makers would have to choose the order in which to adopt the many best practice guidelines on offer, based presumably on relative efficiency. Furthermore, given the cost of drawing up costed guidelines, if they are to have only an incremental effect on budgetary reallocation, it may be that the marginal cost of drawing up guidelines will exceed the marginal benefit of their implementation. Finally, with continuous changes in medical technology and prevalence of diseases, guidelines would have to be updated regularly. Many of these criticisms are not intractable, however, and the field deserves the attention of researchers.

1.3.2 Empirical

> *"The broad principle that informs most recent work on capitation funding is that the main yardstick for deciding whether a putative 'needs factor' should be used as a basis for capitation is whether it explains actual spending patterns amongst plans in a statistically significant manner"*
>
> Rice and Smith, (1999:6)

Empirical modelling, often known as risk-adjustment (although the normative approach can also be considered risk-adjustment), has become increasingly popular method of resource allocation in health care. At least 19 health care systems employ risk-adjustment to allocate significant amounts of their health care budget (Rice and Smith, 1999).

In the UK, the first weighted capitation formulae were hybrids of the normative approach as described above (Department of Health and Social Security, 1976) but this has gradually given way to an empirical approach such as Carr-Hill et al. (1994), which has now been adopted in all four countries in the UK for much of their health care budgets. The UK approach uses electoral ward data from the Census as its principal unit of analysis, with synthetic GP practices used for GP budgets. Ward-level use of health services is standardised for age and gender and this variable is

regressed on a number of measures of need based on socio-economic, environmental and morbidity variables from the Census, as well as variables relating to access to health services and supply-side characteristics. Access and supply-side variables, as well as need variables with 'counter-intuitive' signs, are controlled for in estimating an electoral ward's need for health care resources (Carr-Hill et al., 1994; Scottish Executive, 1999). This small area approach has also been adopted in Sweden (Andersson et al., 2000) and Finland.

Elsewhere in Europe and in the US, individual-level data is more often used for empirical risk-adjustment. In these cases, age and gender are modelled as covariates, rather than purging the utilisation variable of their effects and using this variable as the response variable. In many US studies, age and gender are supplemented only with variables indicating previous inpatient or outpatient diagnoses, or indicators of chronic illness from prescribing data, as covariates. Many European studies do not include diagnostic information due to data constraints, but do include socio-economic variables such as indicators of urban residence and receipt of disability payments. Van de Ven and Ellis (2000), Rice and Smith (1999) and Newhouse (1998) offer comprehensive reviews.

The advantages of the empirical approach are that it provides a direct measure of the relationship between need and use of health care. Second, data on use of services is usually routinely available. Moreover, these datasets tend to be very large, running to several million observations for some US studies (Hornbrook et al., 2001; Deb and Burgess, 2002). The disadvantage of the empirical approach is that need is inferred from contact with the health service, so that if certain groups are systematically excluded from consuming their 'fair share' of health services, the approach can provide a biased estimate of need. However, a number of adjustments have been proposed to overcome this 'unmet need' problem.

In conclusion, since the normative approach does not offer a direct link between need and use of services, and given the lack of data in Ireland for many of the applications of the normative approach reviewed above, we choose to use the empirical approach in this study. Therefore, further details on previous research using the empirical approach are provided in Chapter 2.

1.4 CONCLUSIONS

The purpose of this study is to assess variations in prescribing expenditure amongst recipients of GMS services in Ireland, with the intention of informing the GMS prescribing budget-setting methodology. The structure of the thesis was described. Next, we outlined the GMS scheme and its current budget setting formula. We found that it followed a risk-adjustment approach. We then examined ways of setting health care budgets based on risk-adjusted capitation formulae. Hybrids of the normative approach have been applied in a number of health care systems, most notably in the UK and Sweden. The principal advantages of the normative approach are its relative simplicity and the fact that it measures need directly, rather than using utilisation as a measure of need, which might be contaminated by access to health services and variations in the quality of those services. However, its principal disadvantages are that arbitrary resource weights are assigned to levels of need; the measure of need that is adopted may be the 'wrong' one, and the approach can be data intensive. The empirical approach is much more commonly applied. The next chapter will address additional issues in empirical risk-adjustment modelling.

2. PREVIOUS RESEARCH ON RISK-ADJUSTMENT AND RELATED UTILISATION FUNCTIONS

2.1 INTRODUCTION

This wide-ranging chapter has two primary concerns, divided into two main sections. In section 2.2, we examine the specification of a consistent and precise model of health care utilisation for empirical risk-adjustment. In section 2.3, we examine the use of utilisation models for budget setting.

The specification of a health care utilisation function generates a number of conceptual and empirical challenges. We find that the 'standard' empirical risk-adjustment model usually considers only health needs variables in the health care utilisation function. The exclusion of variables that affect utilisation but are not measures of need has been criticised by Schokkaert and van de Voorde (2000) among others. Consequently, we broaden the review of previous research to include relevant theoretical and empirical studies of health care utilisation from other areas of health economics, not just risk-adjustment. The objectives of the review are to determine what covariates should be included, as well as each covariate's expected sign and approximate order of magnitude; how the response variable should be specified and any econometric issues.

We begin with a review of the theoretical models of health care utilisation. These focus either on the role of the individual or of the physician as the key decision-maker. The former include models such as Grossman (1972), while the latter are used for examining the effects of physician practice style or supplier-induced demand (McGuire, 2000 for instance). Next, we supplement these theoretical models of utilisation with a review of empirical studies of utilisation. We categorise the variables included in utilisation studies into five types – demographic, socio-economic, health status, access to health care and physician characteristics. The measurement of health status is one of the more conceptually challenging issues in the field and we consider it in some detail. Equally, the measurement of physician practice style is problematic and we propose a number of new measures of it, based on the physician's aversion, or otherwise, to prescribing. Based on the theoretical and empirical literature, we then draw conclusions on the covariates to include in the

utilisation function and their expected effects, as well as on the specification of the response variable.

Next we examine econometric issues. Health care utilisation data usually have a number of features that pose econometric difficulties, including a large proportion of non-users of health care, a positively skewed utilisation distribution and a number of very high expenditure patients. We review approaches to solving these difficulties, as well as other econometric considerations, including the use of a hierarchical dataset and potentially endogenous covariates. Thus, section 2.2 offers a review of the principal concerns in specifying a consistent and precise utilisation model.

Section 2.3 examines what needs to be considered when making the step from a utilisation function to a budget-setting formula. A number of criteria for model evaluation are standard in econometrics, such as theoretical consistency, goodness of fit and predictive ability. In order to generate budgets from a utilisation function, two other criteria that we need to consider are distributive consequences and risk-exposure of budget holders. Under distributive consequences, which is reviewed in section 2.3.1, we examine the concept of equity of resource allocations based on each budget-setting formula. One of the criticisms of the empirical approach is that using utilisation as a measure of 'need' may bias resource allocation against particular groups. We examine the 'unmet need' phenomenon and discuss candidate remedies. Finally, risk-adjustment studies consistently find that an individual's health care utilisation is difficult to predict, meaning that the budget holder can be exposed to significant risks. Risk management techniques used on the IDTS and in international health care are reviewed in section 2.3.2.

Thus this chapter draws on previous literature to focus on the key challenges in risk adjustment. We review theoretical and empirical utilisation functions, which focus not just on risk-adjustment but also on demand for health care and physician profiling; we discuss econometric concerns and we describe additional considerations when using a utilisation function for budget setting. Many risk-adjustment studies do not review literature outside of the risk-adjustment field in much detail, except to deal with econometric concerns. The remainder of the

introduction examines why we consider a review of the risk-adjustment literature only to be an inadequate review of previous literature for our purposes.

The empirical approach to risk-adjustment – which we refer to simply as risk-adjustment for the remainder of the study – can be specified as follows:

$$y_i = f(x_i), \tag{2.1}$$

where y_i is health care expenditure on individual i and x_i is set of characteristics of individual i that determine their health care expenditure. In risk-adjustment studies, the response variable usually is 12 months expenditure on health care, either all health care, or one particular type of health care such as primary care or (as is the case in this study) primary care prescribing. In what Schokkaert and van de Voorde (2000) refer to as the 'standard' risk-adjustment model x_i consists of variables that measure health care need only. Other factors that affect utilisation, such as access to services or provider characteristics are not included. Schokkaert and van de Voorde, as well as the UK studies in the tradition of Carr-Hill et al. (1994) distinguish between determinants of health services that are related to an individual's need, x_{ni}, and those that are related to access to services or are not 'legitimate' indicators of need, x_{ai}. Supply-side characteristics, z_j, are also included:

$$y_i = f(x_{ni}, x_{ai}, z_j). \tag{2.2}$$

Variables that may be included in x_a (suppressing the i subscript) are costs of accessing health care and what Carr-Hill et al. (1994) refer to as 'counter-intuitive' signs on health needs variables. For instance, Carr-Hill et al. (1994) consider the variable representing the proportion of black residents in an electoral ward as 'counter-intuitive' because it had a negative sign such that it represented inequity rather than lower health need.

The standard risk-adjustment model excludes x_a variables and z_j variables because they do not measure the need of the individual and an individual's health care capitation should not be affected by non-need variables. Carr-Hill et al. (1994) excludes counter-intuitive x_a variables, while the Scottish Executive (1999), Cole

(2000) and Schokkaert and van de Voorde (2000) argue cogently that both x_a and z_j should remain in the estimation model, but should then be excluded from the generation of budgets. These variables explain systematic differences in utilisation between individuals and to exclude them would be to mis-specify the utilisation model. As Cole (2000: 288) states, *"variables with counter-intuitive signs compensate for the excess effects of other variables in the model, so that excluding them removes this opportunity for negative feedback"*. Schokkaert and van de Voorde (2000) simply view this as good scientific practice, distinguishing between the *explanatory* exercise on one hand – the estimation of a utilisation function – and what they term the *normative* exercise on the other – the determination of legitimate risk-adjusters.

This critique of the standard risk-adjustment model leaves us with two gaps in the literature, however. First, we need to know the expected sign on each x variable, in order to identify variables with 'counter-intuitive' signs. Risk-adjustment studies are primarily concerned with the estimation of unbiased and precise predictions of health care expenditure, rather than explaining the process of generating a particular level of health care expenditure. While schematic models of health care demand are provided in UK studies in the tradition of Carr-Hill et al. (1994), we know of no risk-adjustment study that derives a model of health care utilisation from first principles. As such, the process guiding the allocation of x variables into categories x_a and x_n appear to be intuition and knowledge of previous research. Hence, the first gap in the risk-adjustment literature, given the above critique of the standard model, is the use of theoretical models of health care utilisation to guide hypotheses on expected signs for each variable. Theoretical models of health care utilisation are of two types. First, there are models of an individual's demand for health care, stemming from Grossman (1972). Second, there are models of health care utilisation focusing on physician behaviour. Both these strands of literature refer to health care utilisation in general terms and so are as relevant to prescribing expenditure as other types of health care. They are reviewed in Section 2.2.1 and theoretical predictions are identified.

The second gap in the literature is in variable inclusion. Since the standard risk-adjustment model is concerned with estimating the effects of x_n only, there is a

paucity of risk-adjustment studies that include x_a or z_j variables. In order to undertake a comprehensive review of these variables, we need to examine other empirical literature on utilisation functions. Therefore, section 2.2.2 reviews empirical, multi-variate utilisation functions covering such areas as the demand for health care, profiling of physicians (or types of physician practice such as GP fundholding in the UK or non-profit hospitals in the US) and quality assurance in health care as well as risk-adjustment. Since the theoretical models in section 2.2.1 often refer to concepts such as the stock of human capital, investment in and the depreciation of health and physician effort, section 2.2.2 covers the measurement of these concepts. In addition, a number of empirical studies include other variables that have been shown to have a systematic effect on health care utilisation but are not referred to in the theoretical models. Thus, we conclude section 2.2.2 with a description of the variables included in our risk-adjustment model and the hypothesised effects of these variables. We allow theoretical predictions reviewed in section 2.2.1 to guide our hypotheses where possible. For those variables that are included but on which theoretical models have been silent, we generate a hypothesised effect based on previous empirical literature, if we are satisfied that their empirical models are well-specified.

One element of the specification of model (2.2) remains to be reviewed. The construction of y_i is covered above, while sections 2.2.1 and 2.2.2 cover the variables included in x_n, x_a and z_j and their expected signs. Thus, section 2.2.3 examines the specification of $f(.)$, the functional form. The risk-adjustment literature has applied a number of innovative approaches to specifying the functional form of (2.2). We review this literature, supplemented with other econometric studies on health care utilisation.

In conclusion, this wide-ranging chapter examines the theory of health care utilisation, empirical estimations of health care utilisation functions, and the use of empirical utilisation functions for risk-adjustment.

2.2.1 Theoretical Models of Utilisation

One of the most cited papers in the literature on health care demand is Grossman (1972). His household production approach to health care demand has become known as the Grossman model. This model focuses on the choices made by the individual in determining their health status and their demand for medical services. We review this model in section 2.2.1.1. However, it is widely acknowledged that owing to information deficits, the health care consumer is not sovereign and defers many important consumption decisions to their more informed physician. Thus, we also consider another type of model of health care utilisation, one that models the physician as the chief decision-maker. These are reviewed in section 2.2.1.2. Finally, some conclusions on theoretical models of health care utilisation are drawn in section 2.2.1.3.

2.2.1.1 The Grossman Model

The starting point of many models of demand for health care is the Grossman (1972) model of the demand for health. The following description borrows significantly from Grossman (2000).

The inter-temporal utility function of the typical consumer (U)[2] is as follows:

$$U = U(\phi_t H_t, Z_t) \tag{2.3}$$

where H_t is the stock of health, $\phi_t H_t$ is the consumption of health and Z_t is the consumption of other goods and services.

Health in any period is a function of health in the previous period, depreciation in health since the previous period and investment in health since the previous period as follows:

$$H_t = H_{t-1} - \delta_{t-1} H_{t-1} + I_{t-1} \tag{2.4}$$

where δ_{t-1} is depreciation in health status during period t-1 and I_{t-1} is gross health investment during period t-1. The Grossman model adopts a household production approach, such that the household production function for health includes health care

[2] The individual level subscript *i* has been suppressed throughout Chapter 2 for notational simplicity.

services M_t and health producing time TH_t. The stock of knowledge (or human capital exclusive of health capital) E is treated as exogenous, as follows:

$$I_t = I_t (M_t, TH_t; E) \tag{2.5}$$

For simplicity, the Grossman model treats M_t as a scalar representing health services purchased in the market, although the formulation can be extended to include a vector of other health producing goods and services including housing, diet, alcohol and tobacco consumption. Grossman (2000: 353) points out that M_t can be considered as the primary health input purchased in the market, which may or may not be health care.

The production function for other goods and services depends on a vector of goods and services X_t that contribute to Z_t and time inputs T_t, as well as exogenous E:

$$Z_t = Z_t (X_t, T_t; E) \tag{2.6}$$

In the household production approach, both market goods and time are scarce. The goods budget constraint is the present value of expenditure on goods equated to the present value of income earned over the life cycle plus initial assets:

$$\sum_{t=0}^{n} \frac{P_t M_t + Q_t X_t}{(1+r)^t} = \sum_{t=0}^{n} \frac{W_t TW_t}{(1+r)^t} + A_0 . \tag{2.7}$$

P_t and Q_t are the prices of M_t and X_t, W_t is the hourly wage rate, TW_t is hours of work, A_0 is initial assets and r is the market rate of interest. Meanwhile, the time constraint is as follows:

$$TW_t + TH_t + T_t + TL_t = \Omega. \tag{2.8}$$

Ω is the total time available in any period (= 8760 hours per year if periods are measured as years) and TL_t is time lost from market and nonmarket activities due to illness or injury.

By substituting for hours of work from Equation (2.8) into Equation (2.7) we get the 'full wealth' constraint:

$$\sum_{t=0}^{n} \frac{P_t M_t + Q_t X_t + W_t (TL_t + TH_t + T_t)}{(1+r)^t} = \sum_{t=0}^{n} \frac{W_t \Omega}{(1+r)^t} + A_0 . \tag{2.9}$$

Initial assets plus the present value of earned income if all hours were spent at work (the right hand side of the equation) is equal to spending on market goods, time spent at non-market production and time lost due to illness, all in present value terms (the left hand side of the equation). Equilibrium quantities of H_t and M_t can be found by

maximising the utility function in Equation (2.3) subject to the constraints in Equations (2.4), (2.5) and (2.9), as well as some boundary conditions relating to the time of death, which is endogenous, and the stock of wealth at death.

The first order optimality conditions for gross investment in period t-1 are

$$\frac{\pi_{t-1}}{(1+r)^{t-1}} = \frac{W_t G_t}{(1+r)^t} + \frac{(1-\delta_t)W_{t+1}G_{t+1}}{(1+r)^{t+1}} + \ldots + \frac{(1-\delta_t)\ldots(1-\delta_{n-1})W_n G_n}{(1+r)^n} + \ldots$$
$$\ldots + \frac{Uh_t}{\lambda} G_t + (1-\delta_t)\ldots(1-\delta_{n-1})\frac{Uh_n}{\lambda} G_n \qquad (2.10)$$

$$\pi_{t-1} = \frac{P_{t-1}}{\partial I_{t-1}/\partial M_{t-1}} = \frac{W_{t-1}}{\partial I_{t-1}/\partial TH_{t-1}} \qquad (2.11)$$

In these equations, h_t is the total number of healthy hours; $Uh_t = dU/dh_t$, the marginal utility of healthy time; λ, the marginal utility of wealth; $G_t = dh_t/dH = -(dTL_t/dH_t)$, the marginal product of the stock of health in the production of healthy time and π_{t-1}, the marginal cost of gross investment in period t-1. Equation (2.10) states that the present value of the marginal cost of gross investment in period t-1 equals the present value of marginal benefits of investment. The discounted wage rate plus the discounted monetary value of an increase in utility due to a one unit increase in healthy time (Uh_t/λ) by the marginal product of health capital produces the discounted marginal benefits. Equation (2.11) shows the condition for minimising the cost of producing a given quantity of gross investment. It states that total cost is minimised when the marginal cost of medical care equals the marginal cost of time.

Grossman (1972) empirically tests the 'pure investment' model, that is, the model where healthy time does not of itself enter the utility function. We apply a general production function for healthy time as follows:

$$H_t = 8760 - BH_t^{-C} \qquad (2.12)$$

where B and C are positive constants. Three basic structural equations are proposed:

$$\ln H_t = \varepsilon \ln W_t - \varepsilon \ln \pi_t - \varepsilon \ln \delta_t, \qquad (2.13)$$

$$\ln \delta_t = \ln \delta_0 + \tilde{\delta}_t, \qquad (2.14)$$

$$\ln I_t \equiv \ln H_t + \ln(1 + \tilde{H}_t/\delta_t) = \rho_H E + (1-K)\ln M_t + K \ln TH_t. \qquad (2.15)$$

ε is the elasticity of marginal efficiency of health capital, that is, the percentage change in health stock supplied by a 1% increase in the rate of return on health investment; tilde denotes a percentage time or age derivative; ρ_H is education elasticity of health investment, that is, the percentage increase in gross health investment supplied by a 1% increase in education; and K is the fraction of total cost of gross investment accounted for by own-time.

These three equations and the least-cost equilibrium condition expressed in (2.11) generate the following reduced-form demand curve for medical care:

$$\ln M_t = [(1-K)\varepsilon + K]\ln W_t - [(1-K)\varepsilon + K]\ln P_t + \rho_H(\varepsilon - 1)E + ... \\ ... + \tilde{\delta}(1-\varepsilon)t + (1-\varepsilon)\ln\delta_0 + \ln(1+\tilde{H}_t/\delta_t) \quad (2.16)$$

The demand for medical services is a function of wages, the price of medical care, the stock of human capital, age (t), the rate of depreciation in the initial period and net disinvestments relative to depreciation. The initial rate of depreciation is unobserved, while if net health disinvestments relative to depreciation is small, the last term can be ignored. As such demand for medical services is essentially a function of the first four variables and an unobserved term $u_t = (1\text{-}\varepsilon)\ln\delta_0$. The model predicts that the coefficient on wages is positive and the coefficient on medical care prices is negative. In addition, if $\varepsilon < 1$, the coefficient on education is negative and the coefficient on age is positive. So long as $\ln\delta_0$ is not correlated with the covariates, u_t can be treated as a disturbance term and the above model can be estimated using OLS. Further details, including an overview of the 'pure consumption' model, are available in Grossman (2000).

2.2.1.2 *Physician-focused Models of Utilisation*

Owing to information asymmetries, physicians are at least as important as patients are in making health care consumption decisions. This leads to two questions. First, how does the principal-agent relationship affect health care consumption decisions? Second, given that patients defer important consumption decisions to physicians because of their uncertainty about their health and the appropriate health care to consume, to what extent can physicians eliminate that uncertainty? Moreover, to what extent can physician's performance be monitored, given that uncertainty?

Literature on physician utility functions forms a subset of applied work on agency theory and focus on the role of ethics as an argument in the utility function. Scott and Shiell (1997) suggest that unlike conventional principal-agent literature, the utility functions of the physician and the patient are not independent. The physician's professional ethics mean that the perceived gain in health status of the patient through treatment is included in the physician's utility function. Jaegher and Jaegher (2000) also include professional ethics in the physician's utility function, modelling it as a 'moral cost' preventing excessive supplier-induced-demand. Mechanic (1998) outlines the relationship between ethics and trust highlighting the importance of trust for effective treatment. McGuire (2000) interprets medical ethics in the context of a situation where there are a number of courses of action, the physician chooses the 'medically correct' one. The role of economic incentives is incorporated into this model of ethics by allowing there to remain a number of 'medically correct' choices all of which are ethical, with economic incentives having the potential to affect the final choice. As Hillman (1990) puts it, *"where most physicians will act in the patients' best interest when the medical decision is clear-cut, the effect of financial incentives may be more important in areas where the correct decision is not clear"*. Ma and McGuire (1997) include ethics in a model of physician behaviour, with ethics setting a lower bound on the health benefits that a physicians is willing to provide.

Thus, the role of ethics in the physician's utility function can be framed, as per Scott and Shiell (1997), as :

$$U_d = f(HS_p, Y, X) \qquad (2.18)$$

where U_d is the physician's utility, HS_p is the health status of the patient, Y is physician income and X is a vector of other utility bearing attributes including leisure, professional status and intellectual satisfaction. In terminology used by Gravelle, Dusheiko and Sutton (2002), the physician is a quasi-altruist, concerned with patient welfare as well as their own income and other 'selfish' factors.

We turn now to physician uncertainty in the production of health. Phelps (2000: 238) highlights the physician uncertainty in the perceived efficacy of treatment, described by the shape and slope of the health producing technology $I_t()$ in (2.5) above. Given that there are over 15,000 codes in the International Classification of

Diseases (ICD-9), many argue that physicians cannot diagnose accurately each illness presented before them (see Phelps, 2000 for a review). In addition, they do not know the specific effects of each treatment, that is, they do not know precisely H_t or I_t. Therefore, it is unsurprising that levels of utilisation differ by physician. Some may seek more information through further diagnostic testing, conference with colleagues or referral of the patient to a specialist; others may try some treatment programme; others again might continue to monitor the patient without initiating a treatment programme, while others again might decide that there is nothing that can be done. The physician's choice of course of action depends on their assessment of the costs of collecting the additional information required for a more specific diagnosis and the benefits of collecting that information on the health of the patient[3].

McGuire (2000) highlights three types of uncertainty in the assessment of utilisation of health care. First there is irreducible uncertainty, that is, that neither the patient nor the physician has complete information on the initial health status of the patient - so there is uncertainty in diagnosis - or complete information on the appropriate treatment programme and its exact outcome. In order to model this individual unobservable heterogeneity, we can rewrite (2.5) above as:

$$I_t = I_t (M_t, TH_t, u_t; E) \qquad (2.19)$$

where u_t is a random variable with zero mean and variance of σ_u. Second, there is uncertainty about the 'effort' applied by the physician owing to asymmetric information. In many industries 'effort' or 'quality' is difficult to contract as it is difficult to measure. Measurement difficulties are especially acute in health care, due to the degree of specialist knowledge by the physician, the fact that health care is an 'experience good' (where in order to assess its quality one has to experience the services) and is non-tradable and heterogeneous. Effort can relate to the costs of making an accurate diagnosis or the effective implementation of a treatment programme. Uncertainty in effort can be expressed in the production function as:

$$I_t = I_t (M_t, TH_t, u_t, e_{pt}; E) \qquad (2.20)$$

where e_{pt} is unobservable effort by physician p. Note that the presence of the random variable u_t means that e_{pt} cannot be inferred from health outcomes and therefore remains unobservable.

---

[3] Goldman and Grossman (1978) modified Grossman's original health production function by including an argument for physician quality. The emphasis in consumption choices remained on the

Third, there is uncertainty in physician's skills, again due to asymmetric information. This is equivalent to treating the e_{pt} above not as representing physician effort, but rather as an unalterable characteristic reflecting physician's skill at diagnosis or treatment. Physicians differ in their training, specialist interests and ability. As such, patient's health investments may differ due to differences in physician ability, rather than effort or observable inputs. Note that the random term must remain in the production function, or again physician skill could be inferred from outcomes.

2.2.1.3 Conclusions

Variations in utilisation of health services by individuals that may be due to a number of reasons. First, there is irreducible uncertainty. Second, there are differences among patients – differences in health status, incomes, tastes and preferences including aversion to risk or aversion to the process of health care, and differences in the 'full-price' of health care – which the Grossman model attempts to organise. Third, there is physician-related variation due to differences between physicians in skill, tastes and preferences including effort and risk aversion, and demand inducement. The interaction between patient decision-making and physician decision-making has not been formulated rigorously. As Zweifel and Breyer (1997: 123) state:

> *"[T]he treatment process is governed by the individual in a few aspects only (choice of doctor, preference for drug, termination of treatment), [meaning that] the physician obtains leeway to pursue his own objective. This consideration calls for an investigation of physician decision-making...[i]n a sense, these partial analyses must remain unsatisfactory, however, since they fail to portray the interaction between the physician and his patient. Through comparing and combining empirically tested implications of such partial models, it may still be possible approximately to predict the effects of particular health policies."*

Therefore, we draw on the predictions generated by the Grossman model, while considering the role of the physician, namely, we expect health care utilisation to be positively related to wages and age and negatively related to education and medical prices and we also expect physician effects to matter.

---

individual, however.

2.2.2 Empirical Models of Utilisation: Variable Selection

There are well over one hundred published studies of multi-variate regressions of factors affecting health care utilisation covering many different policy questions, as described in the introduction to Chapter 2. While the choice of covariates for some relate directly to the theoretical models outlined above, others do not. For instance, the standard risk-adjustment model uses measures of individual health care need only, as outlined above. In general we can categorise covariates as relating to demography, health status, socio-economic characteristics, characteristics of the health care market including access and demand inducement, and characteristics of physicians.

Demographic variables include age, gender, household or family composition, religion, ethnicity and urbanisation. Indicators of health status include self-assessed and clinically assessed health, mortality and information drawn from previous use of health services. Socio-economic characteristics include education, income, housing tenure, social class, as well as variables relating to the price of medical services including insurance status and co-insurance rate. Characteristics of the health care market include number of health care workers per capita, or number of health care facilities per capita, as well as distance from the individual to health care providers. These characteristics are used either as variables to test for supplier-induced demand or as measures of access to health care. Finally, as indicators of practice style, physician characteristics are considered.

Table 2.1 exhibits illustrative studies of utilisation functions. They are grouped by type of dataset. First, there are large health care administrative datasets, where there are usually tens or hundreds of thousands of observations, as used in most individual-level risk-adjustment studies. Second, there are household surveys. These are the main source of data for demand for health care studies. Third, there are small, health service site surveys, which collect data in health centres such as outpatient waiting rooms. In general, they collect a more detailed dataset than the large administrative datasets, but on a smaller number of people. However, they are restricted to the population of attendees, and may suffer from sample selection bias if they are used to generalise results across the entire population. Finally, there are studies using small area data, including UK risk-adjustment exercises. The studies included in Table 2.1

are examples of well conducted studies using each type of dataset, although they do not purport to be definitive representations of each type. The text that follows considers not only the studies described in Table 2.1 but also many others. As such Table 2.1 simply sets the scene.

Table 2.1

Model, Estimation Procedure and Variables included in Selected Health Utilisation Studies

			Covariates				
Study	**Model type & Estimation**	**Response variables**	**Demography**	**Health Status**	**Socio-economic**	**Access/Health Care Market**	**Physician Characteristics**
Individual-level Risk-Adjustment Datasets							
Blough et al. (1999)	Risk-adjustment Generalized linear model N = 126,393	Individual medical expenses per annum	Age, Gender, Family composition, Contract status (sponsor, other)	ADG, ACG, CDS, DCG			
Lamers (1999a)	Risk-adjustment OLS N=55,907	Individual health care expenses per annum	Age, Gender, Urbanisation	CDS, Disability payments	Employment		
Lamers (1999b)	Risk-adjustment OLS N=52,674	Individual health care expenses per annum	Age, Gender, Urbanisation	DCG, Health survey, Disability payments	Employment		
Breyer (2001)	Risk-adjustment OLS, tobit N=77,015	Individual health care expenses per annum	Age, Gender, Marital status	Death, Prior health expenditure, Disability pensioner	Contributable income		
Household Surveys							
Cameron et al. (1988)	Demand for Health Care Count models Australian health Survey N=6,539	Number of discrete health care utilisation variables	Age, Gender, Marital status, Ethnicity,	GHQ, Chronic illness, Recent illness, Illness concern, Lost days of activity	Occupation, Education, Income		
Propper (2000)	Demand for health care	Any health care use,	Age, Gender,	Activities of daily living,	Housing, Employment,	Regional indicators of NHS quality	

			Covariates				
Study	Model type & Estimation	Response variables	Demography	Health Status	Socio-economic	Access/Health Care Market	Physician Characteristics
	British Household Panel Survey N=20,598 - 21,002	Inpatient use, Dental use,	Family composition,	Smoker	Income, Industry, Political preferences		
Deb & Trivedi (1999)	Demand for health care RAND Health Insurance Experiment N=2,823	Outpatient use	Age, Gender, Ethnicity	Self-rated health, health concern	Income, Health insurance variables		
Grossman (1972)	Demand for health care National Opinion Research Centre N=7,803	Medical expenditure	Age, Gender, Family size		Income, Education, Wage rate		
Health Service Site Surveys							
Acton (1975)	Demand for health care OLS Random sample of outpatient departments	Outpatient visits, Inpatient days, Private physician care	Age, Gender, Ethnicity, Household size	Health status rating scale	Education, Household wealth, No. of household earners, Medical insurance, Wage earnings, Other income, Welfare assistance, Employment status	Distance to hospital, mode of transport, habit of using the clinic, waiting time	

Study	Model type & Estimation	Response variables	Covariates				
			Demography	Health Status	Socio-economic	Access/Health Care Market	Physician Characteristics
Small Area Data							
Rice (1999)	Risk-Adjustment 2SLS N=349	Cost per Age and Gender standardised prescribing unit	Ethnicity, Religion, Population density, Family composition, Marital status	Standardised mortality ratio, Limiting long-standing illness, Permanent sickness	Census-related socio-economic indicators, including Income support, Employment, Housing tenure, Education	Access to GP, nursing home, acute care, geriatric beds, private hospital beds	Number of partners fundholding patients per GP type of computer system
Reid et al. (1999)	Utilisation function of hospital admissions OLS N=120	Hospital admission rates by General Practice	Age, Gender	Chronic illness	Census-related socio-economic indicators, including income support, employment, housing tenure, education	Distance to hospital	Cervical Smear uptake, Patients per GP, Average age of GPs, No of Partners, Generic Prescribing %, Corticosteroids to bronchodilators, Single handed, Female GP in practice, Fundholding, Manager, Nurse, On minor surgery list, On obstetrics list, On child health surveillance list, Vocational training scheme, Medical students, Standard of premises

NOTE: ADG = Ambulatory Diagnostic Group; ACG = Ambulatory Care Group; CDS = Chronic Disease Score; DCG = Diagnostic Cost Group; GHQ = General Health Questionnaire.

Although all studies outlined in Table 2.1 essentially estimate a utilisation function, there are clear differences in the type for variables applied, depending on the focus of the study and data restrictions. Risk-adjustment studies usually have good demographic data and good measures of health status, but include few or no variables on socio-economic characteristics or characteristics of health care supply, as outlined in the introduction. Studies of demand for health care include good measures of demographic effects, usually include measures of health status – of a different type to risk-adjustment studies – as well a detailed set of measures of socio-economic status and price of health care. Supply-side characteristics again are rarely measured. Health care utilisation studies that collect primary data, such as Acton (1975) provide details on all the principal aspects of demand, as well as indicators of access to health care. In Acton (1975) physician characteristics are not measured, however. Finally, the principal feature of small area studies is that they use Census of Population data to provide a raft of demographic, health status and socio-economic variables. The studies reviewed also provide a number of measures of supply-side characteristics, including physician characteristics. Reid et al. (1999) is a good example of the effect of the focus of the study on the variables collected. In a study of GP referral behaviour, it collects 17 measures of GP characteristics. Meanwhile, the risk-adjustment studies, household survey studies and health site survey study included in Table 2.1 collected none. Finally, although the response variable is always health care utilisation, the way it is measured differs by study. As in this study, risk-adjustment studies use 12 months expenditure as the response variable in general. Some other studies also use expenditiure, but counts of utilisation are also common. There may be a difference between the relationship between covariates and the response variable depending on whether the response variable is measured as expenditure or counts if that covariate is expected to have an effect on the unit cost of utilisation as well as the quantity consumed. Where we detect differences in the effect of a covariate between studies that use expenditure and studies that use counts, we highlight these and focus more on the results of the studies that use expenditure.

2.2.2.1 Demography

Age and gender are included as explanatory variables in almost all utilisation studies and certainly all risk-adjustment studies. Except for the very young and very old,

age is positively related to health care use, as predicted by the Grossman model. A number of alternative ways of specifying the relationship between age and utilisation exist, including age splines, dummies and polynomials. These are discussed in more detail in Chapter 4.

Where gender is significantly associated with utilisation – which is not always the case – being female is usually positively associated with utilisation (Pohlmeier and Ulrich, 1995; Cameron et al., 1987 for physician consultations; Ryan et al., 1999). Although these are well conducted studies, they are demand studies, and the first two related to number of visits, rather than expenditure as in our study. Meanwhile, some evidence for Ireland suggests that females have lower GP visitation rates than males and are associated with having a regular GP for a shorter time (Watson, 1996). In risk-adjustment studies, age and gender are often interacted and modelled as a series of dummy variables. As such, it is not possible to disentangle the independent effects of age and gender. However, some studies include gender as well as these interactions. Blough et al. (1999) finds that being female is negatively associated with utilisation, while Hornbrook and Goodman (1995) find that it is insignificant. The ambiguity in the variable can be summed up by the following justification of the inclusion of a female dummy: *"to test the hypothesis that females might be more efficient producers of health than males (or vice versa)"* Grossman (1972:50).

Like gender, the effect of marital status on health care utilisation is uncertain. The variable is not included in many individual-level risk-adjustment studies. An exception is Breyer (2001), which found that marital status was insignificant. Amongst demand for health care studies, Cameron et al. (1988) found marital status was insignificant, while Pohlmeier and Ulrich (1995) found that being single was negatively associated with GP utilisation and specialist utilisation.

Ethnicity is occasionally included as an explanatory variable, with ethnic minorities usually having higher utilisation rates than the majority group, and different patterns of utilisation, having controlled for health status and socio-economic status (Sutton et al., 1999b). It is included in small-area level risk-adjustment studies such as Carr-Hill et al. (1994).

Religion is used less often. In Northern Ireland, the weighted capitation formula for general practice prescribing expenditure included proportion of the post-code sector that were Roman Catholic. This did not remain in the preferred model, however, owing to statistical insignificance (Rice et al., 1999).

Family or household composition, such as number of adults and children in the household, and contract status (person in household who is first registered by the health plan is known as the sponsor and others are dependants) are included to measure household decision-making and the possibility of intra-household differences in allocation of health care resources (Hornbrook and Goodman, 1995). Grossman (1972), in an empirical estimation of the pure investment version of the Grossman model, included family size as a regressor in order to adjust for the fact that the income variable also included as a regressor refers to family income while the response variable, health care expenditure, referred to the individual.

2.2.2.2 *Indicators of Health Status*

We review two types of measure of health status. First, there are direct measures of health status. Second, there are measures that are generated from information from previous use of health services.

2.2.2.2.1 Direct Measures of Health Status

Direct measures of morbidity or health status have long been used in utilisation functions. They can be placed in four categories. First there are social insurance payments for disability (eligibility for which requires one's disability to be assessed by a physician). Second, there are clinically measured indicators of health. Third, there are self-rated measures of health. Fourth, there is mortality.

Persons in receipt of disability payments have been used in risk-adjustment models in the Netherlands (Lamers, 1999a), Germany (Breyer, 2001) and Belgium (Schokkaert and van de Voorde, 2000) and the US (Kronick et al., 1995). Kronick et al. (1995) found that Medicaid resource utilisation by disabled persons was more predictable than non-disabled for selected US states. Breyer (2001) found that disabled had approximately twice as much health care expenditure as non-disabled. Van de Ven and Ellis (2000) found a similar magnitude of effect for disabled and

functionally impaired persons in their review of risk-adjustment. Schokkaert and van de Voorde (2000) distinguish between those in receipt of disability benefits for more than one year and less than one year, finding that the more than one year group had 12% higher expenditure than the less than one year group. In addition, both groups had expenditure that was well above average.

The use of clinical factors such as blood pressure has been considered, but are expensive variables to collect. Newhouse et al. (1989) included both dichotomous and continuous physiological health variables including the presence of hypertension, diabetes, anaemia and hernia. These improved the explanatory power of the model over the basic model containing demographic and some socio-economic variables from 1.6% of total variation in expenditure to 4.5% of variation in expenditure. Low birthweight has been used in the British NHS's weighted capitation models, including Carr-Hill (1994), Scottish Executive (1999) and Rice (1999), which are based on small area analysis. This variable may indicate maternal ill health and maternal smoking, as well as child ill health. It is usually rejected from the final UK models due to statistical insignificance. In summary, clinical variables are unlikely to provide promising avenues for the development of risk-adjustment (McCarthy et al., 1995).

Self-rated health surveys are a common feature of the type of household survey that are used in studies of the demand for health care. Propper (2000) uses questions on activities of daily living and smoking status as measures of health status in an assessment of the demand for private health care in the UK based on the British Household Panel Survey.

Owing to the cost of collecting this type of survey data universally, it has rarely been applied in risk-adjustment studies. Newhouse et al. (1989) tested models which included functional health status variables, including general health perceptions, mental health and self-reported chronic illness. These variables performed poorly, however. It was estimated that at most only two to three percent of future individual expenditure variation could be captured with these data. Hornbook and Goodman

(1996) included the RAND-36 health survey[4] in a risk-adjustment model. The inclusion of this indicator improved R^2 from 1.2% for an age-gender model to 4.6%, a slightly larger gain than that reported by Newhouse et al. (1989), but nevertheless quite modest given the costs of collecting the data. As with clinically measured factors, self-reported health status does not seem to be a promising way of predicting future health expenditure for risk-adjustment.

Census-derived measures of ill health have been used with more success in small area risk-adjustment (see for example Scottish Executive, 1999). These measures include standardised permanent sickness ratios and standardised illness ratios. The standardised permanent sickness ratio is based on the respondent's economic activity in the previous week, one response to which is 'unable to work due to permanent sickness or disability'. The proportion of the district electoral division's (DED) population in this category is then age and gender standardised to derive the standardised permanent sickness ratio. Standardised illness ratios are based on a question on the presence of limiting long-standing illness. The biases in these data are outlined under the normative approach above, namely that they are a better measure of physical health than mental health, and that they depend on age, gender and how recently one last visited the GP.

Finally, mortality can be used as measure of health status, either in individual-level analysis, or as a measure of community morbidity in small-area studies, as community premature mortality rates are strongly correlated with community morbidity.

The closer one is to death, the lower one's health status and the greater one's health care expenditure. Van Vliet and Lamers (1998) found that death was associated with 15.3 times more expenditure on health care than survival. Breyer (2001) used a proxy for death in a risk-adjustment model for one German sickness plan, finding it to be highly significant and indicated considerable excess costs associated with death.

[4] The RAND-36 survey is similar to the more well-known SF-36 survey, differing only in how responses are scored.

However, van Vliet and Lamers (1998) highlight problems with the data relating to reliability, validity, availability, manipulation and privacy. They argue that death should not be used as a risk-adjuster because the excess costs associated with death are unpredictable. In their most refined model, which includes a death variable, the expenditures of those who die were still 250% greater than predicted expenditures. On the other hand, Beck and Zweifel (1998) and Breyer (2001) suggest that health plans should be paid retrospectively for the excess costs of death.

While the use of death data is still controversial in individual-level risk-adjustment, small area-level risk-adjustment models justify the use of community premature mortality or mortality as indicators of community morbidity. Van de Ven and Ellis (2000: 806) report the use of 'average number of deaths per 1000 enrolees in prior years at the health plan level' as a risk-adjuster in Belgium. Weighted capitation models in the UK have long used the 0-64 standardised mortality ratio (SMR064), both in normative weighted capitation models (Department of Health and Social Security, 1976; Department of Health, 1988) and empirical ones (Carr-Hill et al. 1994, for example). However, while the degree to which death represents a plausible risk-adjustment factor is reasonably well established with respect to conditions from which you can die, it is less useful for conditions which do not affect life expectancy significantly such as mental illness, as well as areas such as obstetrics and gynaecology.

2.2.2.2.2 *Previous Use of Health Services*

According to Newhouse (1998: 124), since age and gender are not powerful predictors of health care utilisation, explaining between 1% and 3% of total variation, the most promising development in risk-adjustment is likely to be previous diagnosis. The two principal alternative diagnosis-based risk-adjusters are Ambulatory Care Groups (ACGs) and Diagnostic Cost Groups (DCGs). There are a number of other models, including the Disability Payments Scheme (DPS) which is similar in design to the DCG model but has been used principally in disabled populations; a model designed by Hornbrook et al. (1991) based on ambulatory and pharmacy data and the Costly Diagnostic Groups model, used in risk-adjustment in Sweden (Andersson et al., 2000). Other uses of previous utilisation is the direct use

of previous expenditure as a risk-adjuster. Finally, prescribing information has been used to elicit health status.

The key features of all diagnosis-based models are that they place each of the International Classification of Diseases (ICD9) codes - of which there are over 15,000 – into more aggregated diagnostic groups, based principally on clinical criteria, although economic criteria are important as well. Individuals are then assigned to particular categories, depending on their combination of diagnostic groups.

ACGs were designed by Starfield et al. (1991). Each ICD9 code is placed in one of 32 categories, known as Ambulatory Diagnostic Groups (ADGs). Criteria for placement including duration of episode of care, severity of condition, diagnostic certainty, aetiology and specialty of care. A patient is then assigned to one of 93 ACGs based on their particular combination of ADGs, their age and gender. People in the one ACG are expected to have experienced a similar morbidity and resource use over the course of a year (http://acg.jhsph.edu/what/what.html accessed 19 May 2002).

As the name suggests, ACGs originally used only ambulatory diagnoses, although they have since been modified to include inpatient diagnosis, based on the Diagnosis Related Groups (DRG) methodology of assigning inpatient diagnosis, or following a so-called 'hospital dominant' methodology, where inpatient diagnosis is only coded if more than 50% of patients with that diagnosis were hospitalised (van de Ven and Ellis, 2000). Applications of ACGs in risk-adjustment include the Minneapolis Buyers Health Care Action Group (Dunn, 1998), Washington state (Madden et al., 2000) and the HMO market (Hornbrook and Goodman, 1995).

DCGs, developed by Ash et al. (1989), place all diagnoses into one of 543 diagnostic groups. These are further grouped into 118 Hierarchical Condition Categories (HCCs). These HCCs are then used in regressions on medical expenditure to determine expenditure weights. Unlike the ACGs, DCGs have a more inpatient focus. Unwanted incentive effects, such as an incentive to hospitalise discretionary cases, are dealt with by assigning diagnoses for which unwanted incentives are likely

to accrue to an HCC that is then excluded from the expenditure regression (van de Ven and Ellis, 2000).

There have been a number of risk-adjustment studies in the Netherlands that have included DCGs and modified DCGs. Lamers (1999b) extended the basic demographic model used in the Netherlands to include dummy variables identifying if an individual was admitted to hospital in the previous 12 months into any of five DCGs. A model incorporating DCG data for the previous three years was also estimated, the rationale being that not only will inclusion in a particular DCG predict expenditures for the subsequent year, but it will also predict expenditures for further subsequent years, with decreasing accuracy. It was found that the predictive accuracy of the three year DCG model was 8% while it was 6.48% for the one year DCG model and 3.78% for the basic demographic model.

Other uses of diagnostic information in risk-adjustment include models applied in Sweden and New York. The hospital model applied in Sweden developed a separate matrix for the sickest 5% of the population, who consume 50% of total expenditure, using a 'costly diagnostic group' approach, based on hospital admission diagnosis over a specified period. Groups include cancer, ischaemic heart disease, cerebrovascular disease, arthritis, arthrosis, hip fracture, schizophrenia and other psychoses (Andersson et al., 2000). New York state identified high cost groups by creating a series of risk pools defined by geographical region and medical condition. Those medical conditions or procedures which are identified are transplants (bone marrow, heart, liver, lung or pancreas), very low birthweight babies, people who are HIV-positive with low CD4 cell counts and chronic, ventilator-dependent patients (McCarthy et al., 1995).

An additional application of information on previous utilisation is to use previous expenditure as a risk adjuster. Van Barneveld et al. (1997) found that for catastrophic care (nursing homes, mental illness etc) in the Netherlands, the explanatory power of models improved from an R^2 of 0.04 for the basic demographic model to up to 0.51 with the inclusion of multi-year prior expenditures. Breyer (2001) found that including the prior year's expenditure increased R^2 from 11.8% for a demographic model to 37.9%. However, the coefficient on prior year's expenses

means that each health plan would effectively be reimbursed a fraction of the previous year's expenditure. As such, including such a variable shifts risk-adjustment away from full capitation and towards partial capitation.

Previous use of prescribing information can also be used to measure health status, most commonly using the chronic disease score (CDS) methodology. The original CDS (Von Korff et al., 1992) was created by identifying chronic diseases using pharmacy claims. A revised CDS (Clarke et al., 1995), covering a wider range of medicines, identifies a prescription for a particular medication or medication class as indicating the presence of one of 28 chronic conditions.

There have been a number of applications of this approach in risk-adjustment. Lamers (1999a) clustered these 28 conditions into six Pharmacy Cost Groups (PCGs) based on correlations in the conditions' future expenditures. The six-condition model had a similar level of predictive accuracy as the 28-condition model. Lamers and van Vliet (2001) chose conditions both on the basis of predictive ability and incentives. Twenty-two pharmacy cost groups were identified. Assignment to a chronic condition was on the basis of having more than 181 prescribed daily doses of a particular medicine a year, with each individual assigned one chronic condition only. In order to reduce perverse incentives those conditions with the lowest follow-up costs were dropped, reducing the number of groups to 14. This model explained 8.9% of health care expenditures, compared with 5% for the 'demographic' model, which included age, gender, reason for insurance and urbanisation. Hornbrook et al. (2001) use chronic disease scores for risk-adjustment on HMO populations in the US. The inclusion criteria were much more relaxed. If someone had received any prescription of a drug identified as a chronic illness drug, then they were classified as having that illness. This may include incidental users, as discussed below. Finally, Fishman and Shay (1999) developed paediatric chronic disease scores, which had a prediction R^2 of 6.1% versus only 0.15% for a demographic model.

Although information on health status from previous utilisation is clearly a better predictor of health service use than age, gender or even many direct measures of health status, it has been criticised on two fronts. First, it may generate a perverse incentive structure. Second, the extent to which it is an unbiased measure of health

status depends on the role of access to health care and quality of that health care. We discuss both these criticisms in turn.

Where the recipients of finance are responsible for submitting the data on which the capitation system is based, perverse incentives can arise. For instance, providers could be rewarded for prescribing a particular drug by more than the cost to them of prescribing the drug, leading to inappropriate prescribing. If the cost of prescribing someone enough anti-depressants to categorise them as suffering from a psychiatric illness is less than the increase in the capitation budget received for someone that has a psychiatric illness, then there is the potential to either prescribe anti-depressants to the marginal patient, or to report the prescription of anti-depressants dishonestly.

In the case of diagnosis-based variables, efforts have been made to reduce the incentive to hospitalise patients in order to ensure that they are coded in favourable groups. In the DCG model, diagnoses for which there is considerable discretion in admission to hospital are coded to a group that is not included in the re-imbursement mechanism. In the ACG model, the focus in coding is on ambulatory diagnosis principally. To the extent that inpatient diagnosis is included, the 'hospital dominant' algorithm ensures that only diagnoses for which the majority of patients require hospital admission are included. Lamers and van Vliet's (2001) model for the Netherlands only reimburses the 14 conditions for which follow-up costs are highest, providing an incentive for providers to efficiently manage patients with conditions that have low follow-up costs.

Lamers and van Vliet (2001) address the potential incentive to prescribe a particular drug purely to indicate a patient as high cost leading to increased future payments. Under their scheme, a patient is only indicated as having a chronic illness if they have more than 181 defined daily doses per year of a drug that is indicated as a chronic illness drug, which would be difficult to do dishonestly. By increasing the cost to the prescriber of wrongly indicating someone as having a chronic illness, the perverse incentive is likely to be minimised.

Previous use of health services will only measure the underlying concept of health if certain conditions hold. While Lamers (1999a:828) states that *"[p]rescribed drugs*

capture underlying disease validated by a doctor's drug order", there are a number of assumptions required for the exercise to be valid. First, the person suffering the chronic condition must present to their GP. This will depend on patient preferences including the disutility associated with the process of health care, their understanding of illness and perceived and actual availability of GP services. Second the GP must correctly diagnose the chronic condition and write the appropriate prescription, including correct dose. This will depend on GP training, prevalence of the condition, GP attitude to prescribing, and access to secondary care, which is often required for an accurate diagnosis. As reviewed above, Phelps (2000) reported the high variation in utilisation that can be ascribed to physician practice style. The patient must then cash in the prescription. In addition, many drugs have multiple indications, making it difficult to ascribe them to a particular condition. For instance, beta-adrenergic blockers are occasionally used by people suffering anxiety, rather than suffering from hypertension or heart disease, so it is difficult to ascribe them to one condition only. Finally, the system querying prescription data must be able to distinguish between incidental users of a particular drug and sufferers of the condition. For instance, Lamers model found that the prevalence of 'pain and inflammation' was 19.2% of the population of people suffering a chronic condition when the inclusion criterion was one or more prescriptions per annum, but it fell to 2.6% when the criterion was four or more prescriptions per annum. Clearly incidental users form a significant proportion of those in receipt of medicine for pain and inflammation. A similar set of problems arise when using diagnosis-based data such as ACGs as measures of health status. Indeed, Ellis (2002) reports that diagnosis data often under-represent the health status of people with expensive chronic conditions.

In order to deal with the problem of multiple indications of a medicine, the chronic disease scores are relatively conservative in their inclusion criteria, that is, only medicines that are prescribed almost exclusively for that condition are included. There are exceptions however, such as the beta-adrenergic blockers, as described above. Indeed, the indicators of cardio-vascular disease are probably the most controversial of all disease areas, as a number of these indicators refer to the presence of a risk-factor, such as high blood pressure or high cholesterol, rather than the presence of cardio-vascular disease *per se*. However, problems with the definition hamper epidemiological measures of cardio-vascular disease as well.

In order to distinguish between incidental users and those who are chronically ill, one approach is to only identify a chronic illness if a certain amount of a medicine has been prescribed in one year. Lamers (1999a) identifies someone as having a chronic illness if they have four or more prescriptions of the relevant medicine a year. Hornbrook et al. (2001) indicate anyone who has one or more, so their inclusion criteria are much more lax. Meanwhile, the weighted capitation formula for the Netherlands, to be implemented in 2002, identifies a chronic illness if more than 181 defined daily doses (DDDs) have been prescribed in one year (Lamers and van Vliet, 2001). The defined daily dose is an approach to comparing medicine utilisation across individuals by defining what the average dose for the average adult for each medicine is likely to be. Therefore, the Netherlands formula identifies a sufferer of a chronic illness if they have been prescribed what the average adult would get for almost half a year.

Ellis (2002) develops a theoretical model focusing on optimal risk-adjustment when diagnoses are uneven signals of underlying health status. Variations in signal quality are assumed to be due to variations in diagnostic ability, similar to the practice style variations literature described in Section 2.1.1.2, rather than responses to the risk-adjustment mechanism. This is similar to the Irish case, where diagnosis is not used in risk-adjustment, so variations in diagnosis are not due to responses to the payment mechanism. Ellis' (2002) model specifies that true health status can be either high cost or low cost and signalled health status can also be either high cost or low cost. The model predicts that payment weights should be increased if the fraction of the population with a particular diagnosis is small and if variation in signal quality across health plans is low. It also suggests that imperfect diagnosis is not a significant cause for concern in risk-adjustment. We return in Chapter 3 to this problem.

2.2.2.3 Socio-economic Characteristics

The link between socioeconomic circumstances and health care need is well established (Department of Health and Social Security, 1980; Department of Health, 1999). Factors that are associated with additional morbidity include unemployment, poor housing, risky occupations, low social capital and stress induced by multiple

deprivation, while Grossman (1972) focuses on the role of human capital in the efficient production of health status.

Most of the US risk-adjustment models using individual-level data concentrate on demographic and diagnostic or health status variables. Few include measures of socio-economic status. The basic Dutch demographic model uses employment status, based on data on reason for compulsory insurance (Lamers, 1999a; van Barneveld et al., 1998). Breyer (2001) uses contributable income in a German context. The addition of these types of variables generally add little to explained variance. For instance, the addition of contributable income and marital status – which was statistically insignificant – increased R^2 from 11.8% to 12.5% in Breyer (2001).

The UK-based risk-adjustment studies have been among the most progressive in incorporating socioeconomic variables. Their small area methodology allows a raft of Census-derived indicators to be tested for their relationship with use of services. For instance, the prescribing model for Northern Ireland included the 'percentage of persons aged eighteen and over with some qualification' as a needs indicator (Rice, 1999:4). Meanwhile the Scottish prescribing model included such socioeconomic variables as 'percentage of persons aged less than 65 on income support' and 'percentage of persons in manual social classes' (Scottish Executive, 1999:83).

Studies of demand for health care using household surveys or primary data routinely include socio-economic explanatory variables, including income or some measure of household affluence, education, occupation or employment status and housing (Propper, 2000; Cameron et al., 1988). Additional socio-economic variables that are included in demand studies are those that relate to the price of services, such as insurance status, co-insurance rate and fees (Acton, 1975; Manning et al., 1987; Grossman, 1972; Goldman and Grossman, 1978).

In summary, when socio-economic variables have been included in risk-adjustment studies their effects have been modest. Indeed, most US studies do not collect these data at all, we suspect because of their lack of economic significance. They have been more productively used in demand studies.

2.2.2.4 *Indicators of Supply*

The only individual-level risk-adjustment study that includes supply-side measures is Schokkaert and van de Voorde (2000), as far as we are aware. Consequently, this review concentrates on other types of utilisation study. We review two types of indicator of supply. First, there is the quantity of health care available in a market, which is used to measure access to health care or level of market competition, depending on the focus of the study. Second, there are indicators of practice style.

2.2.2.4.1 Access to Health Care / Market Competition

Many studies include variables reflecting access to health care or competition in the market for physician services (usually demand studies to test the inducement hypothesis). These variables include physician density measures such as number of physicians per 1000 population (Richardson, 1981; Cutler and Sheiner, 1999; Coulson and Stuart, 1995; Pohlmeier and Ulrich, 1995); ratio of primary care physicians to specialists (Richardson, 1981; Cutler and Sheiner, 1999); density of other health services such as hospital beds, pharmacies, nursing home beds (Cutler and Sheiner, 1999; Coulson and Stuart, 1995); waiting times (Propper, 2000); distance (Geil, 1997) and density of health services weighted by distance (Carr-Hill et al., 1994; Rice et al., 2000; Scottish Executive, 1999; Haynes, 1999). Since inducement effects are not the focus of our study, and are unlikely to occur in a fully capitated system such as the GMS, we focus on measures of access to health care services.

There is some evidence to suggest that access to health services is an important determinant in health service use. Using a small area methodology similar to that adopted in Carr-Hill et al. (1994), Haynes et al. (1999) estimated the effects of distance to hospital and GP surgery on inpatient utilisation, while controlling for needs and other supply factors. Distance to hospital was measured by linear distance from the population-weighted centroid of each electoral ward to the nearest hospital, while distance to GP surgery was measured as distance from the population-weighted centroid of enumeration districts to the nearest GP surgery. Variables used to adjust for variations in health needs included age, gender and those socio-economic indicators that were significant needs variables in the York weighted capitation

models for acute, geriatric and psychiatric care in England and Wales (Carr-Hill et al., 1994). The study found that each 1km increase in distance to the GP surgery led to a 0.95% fall in acute admissions, and a 1.65% fall in psychiatric admissions, while distance to hospital led to a reduced rate of admission of 0.47% per km for acute and 1.29% per km for psychiatric. At the extreme, there was a 15% reduction in acute admissions from the electoral district closest to its GP surgery to that furthest away, while there was a 37% drop in psychiatric admissions from the electoral ward closest to its GP surgery to that furthest away. The study suggests that the effects of distance on hospital utilisation are considerable, especially for sufferers of psychiatric conditions.

Crucial to the question of accessibility is the measure used for distance. Haynes et al. (1999:427) argued that since they were assuming that all points in each electoral ward were equidistant from the nearest hospital, that the further simplification of linear distance was *"as indicative as road distances"*. Meanwhile, the measure of access to general practice used in UK weighted capitation studies such as Carr-Hill et al. (1994) is essentially number of GPs per head weighted for distance, as follows:

$$A_i = \sum_d \left(\frac{G_d f(d_{id})}{\sum_i P_i f(d_{id})} \right),$$

where A_i is the access score of electoral ward i, G_d is number of GPs in surgery d, d_{id} is the linear distance between population weighted centroid of electoral ward i and surgery d, P_i is the relevant population in electoral ward i and f is a distance deterrence function. Carr-Hill et al. (1994) included an 'intrazonal constant' of ten kilometres, in order that electoral wards which are very close to the GP's surgery do not get ridiculously high access scores. The deterrence function used in Carr-Hill et al. (1994) was the inverse square, although other specifications are possible, and their choice of deterrence function and intrazonal constant are somewhat arbitrary.

In individual-level studies of utilisation, the inclusion of distance as an explanatory variable is rare. Geil et al. (1997) include a dummy for living more than 5 miles from the health facility. Acton (1975) measures miles travelled to the health facility. Probably the most comprehensive treatment of distance is Ryan et al. (1999). This study made a number of estimates of the access elasticity of demand for GP services,

including self-reported estimates of distance, time taken and goods sacrificed to attend GP, as well as postcode sector level estimates of distance and time costs. Interestingly, they found that models based on postcode sector level data on distance and time costs were a better fit and more plausible than individual level data. While they sacrifice within-area variability in distance and time, they offer reduced measurement error and suffer less from endogeneity caused by distance being positively related to health status, with sicker people choosing to live in easier access of their GP. Self-reported distances were not significant as a determinant of GP utilisation, while aggregate level predicted distances were only significant when mode of transport was included. Aggregate level predicted time models, incorporating economic status, found significant deterrent effects of time. Students had the highest time prices, followed by full-time employed, unemployed, part-time employed, home-makers and the retired or semi-retired. The results indicate that a model of aggregate level time prices for various economic groups produces the most satisfactory measure of the access elasticity of demand for GP services.

2.2.2.4.2 Practice Style

Physician-related variations in utilisation only assist us in estimating the factors affecting variations in prescribing utilisation if physicians differ systematically, as is suggested in section 2.2.1.2. There is now an enormous body of evidence, mostly in the medical literature, showing systematic physician-related differences in health care utilisation. These are often estimated using small area or regional variations in utilisation, while controlling for patient characteristics. Phelps (2000), Phelps and Mooney (1993) and Foster and Stano (1990) provide thorough reviews in a US context. Reid et al. (1999) reference the principal UK studies.

The magnitude of the physician-related effects is large in comparison to other factors affecting variations. Phelps (2000: 241) showed that the coefficient of variation among New York counties in 1987 was as high as 0.61 for dental extractions and restorations, 0.48 for paediatric pneumonia, 0.46 for extracranial vascular procedures and 0.42 for depressive neurosis. By comparison, the maximum coefficients of variation due to income and price effects were estimated at 0.02 and 0.04 on average respectively.

There is some evidence that physician-related differences are due to differences in 'practice style'. Connell et al. (1984) found that areas with high admission rates for diabetic care had lower than average level of urgency for admitted cases, and admitted cases had lower than average levels of diagnostic testing. Wennberg et al. (1977) found a significant reduction in the variation in tonsillectomy rates in New England once a feedback and review mechanism was introduced. These studies suggest that the referring physician's beliefs about diagnosis and appropriate treatment programme led to different rates of utilisation for similar levels of morbidity. However, a number of other studies found no differences in the level of 'inappropriate' use between high use areas and low use areas (Roos et al., 1977; Roos and Roos, 1981).

Observed variations cannot be explained by aggregation from individual physician to regional level. Using individual physician data Phelps (2000) found that differences in utilisation from high use physicians to low use physicians were over 2:1. Phelps (2000) also rejects explanations of the observed variations that can be explained fully by differences in such factors as prevalence of illness, demand inducement by physicians, substitution between different treatments or patient preferences for treatment. He concludes that practice style is the principal factor explaining variations in utilisation in health care. Meanwhile, Wennberg and Gittelsohn (1982: 120) state that *"in the absence of general agreement on their value for individual patients the style of practice of individual physician appears to take precedence"*.

While the above literature shows that physician factors significantly affect health care utilisation, those studies that have attempted to model aspects of the physician's practice style have had uninspiring results in general. This review is organised into non-UK studies, UK primary care studies and UK primary care studies concerned specifically with prescribing behaviour.

We review eight non-UK studies that include measures of practice style. Grytten and Sorensen (2001), in an examination of supplier-induced demand, include as physician characteristic control variables age, gender, possession of a specialist degree in community medicine, number of years working in the present practice and working in a group practice. An assessment of mean number of laboratory tests per

consultation by physician found that only gender was significant, while an assessment of the proportion of consultations lasting more than 20 minutes during a normal week per physician found that only being a specialist in community medicine and number of years in the present practice were significant. However, Sorensen and Grytten (1999) found that age and gender have strong effects on rates of consultation. Goldman and Grossman (1978) in a hedonic physician fee equation included physician experience and experience squared; specialty dummies and certification dummies; interaction between experience and specialty dummies; a dummy for membership of a medical school faculty; dummies for education; a dummy for physicians who did not talk to mothers in her own language (it was a paediatric study). Only experience, experience squared and specialty dummies were significant.

Gaynor and Polachek (1995) in a study examining determinants of physician fees, included experience of the physician, as well as dummies for subspecialty, gender and foreign medical school of graduation. Again, experience and subspecialty were significant. Kenkel (1990) uses age, board certification status and specialty to measure physician characteristics. None were significant.

Wilensky and Rossiter (1981) used age, level of outside income, a practice nurse dummy and a dummy for high nurse wages as measures of 'practice style' and found that only age and high nurse wages were significant in explaining primary care utilisation. Cameron et al. (1988) differentiated between GP, specialist physician and hospital physician. However, these variables were not used in the final specification because they added little explanatory power. Schokkaert and van de Voorde (2000) found that GP loyalty was negatively and significantly associated with health care utilisation. It is unclear, whether GP loyalty measures quality of the GP, patient preferences or a combination of the two.

Three UK primary care studies that include measures of practice style are included. Hancock et al. (1991), in a small-area analysis of health service utilisation, used proportion of GPs over 65 as a 'style' variable. However, it was removed in a backward stepwise regression, owing to insignificance. Reid et al., (1999) included a large number of indicators of practice style in an analysis of referral patterns as

described in Table 2.1. These included cervical smear uptake; mean physician age in the practice; number of partners in the practice; panel size; percentage of generic prescribing; ratio of corticosteroids to bronchodilators; dummy for fundholder; dummy for female physician in the practice; dummy for practice manager; dummy for practice nurse; dummies for minor surgery, obstetrics, child health surveillance, vocational training, teaching medical students; a rating of the premises. Three of these variables were statistically significant but explained only a tiny proportion of additional explained variance. These were cervical smear uptake, child health surveillance and minor surgery. All are supposed indicators of quality and all were positively associated with referrals, suggesting (unexpectedly) that higher quality GPs refer more patients to hospital. Gravelle et al. (2002) included a vector of practice characteristics in an analysis of practice level admission rates to hospital. These included average age of GPs, gender balance, panel size and practice opening times. They found that these variables were both individually insignificant and jointly insignificant.

Three UK studies of prescribing style included measures of GP practice style or other such quality indicators. Morton-Jones and Pringle (1993), in a comparison of prescribing in dispensing and non-dispensing practices, used a dummy for having an appointment system and percentage of generic drugs prescribed as measures of style or quality. The latter was significant.

Wilson et al. (1996), in an examination of prescribing cost differences between fundholding and non-fundholding practices, included a dummy for training practice and number of partners as measures of style or quality and found that both were significant. Whynes et al. (1997), in an examination of variations in prescribing costs, included as practice style effects number of branch surgeries; number of patients per GP; management hours per GP; non-medical hours per GP; whether it is a training practice; percentage of generic prescribing and dummies for training practice, use of a practice formulary and fundholding. Only use of a practice formulary, fundholding and proportion of generic prescribing were statistically – and negatively in each case – related to prescribing costs. Whynes et al. (1995), in a comparison of fundholding and non-fundholding practices, used the above variables and nurse hours per GP; dummies for presence of a formal practice agreement,

computer use for the prescribing budget, computer use for repeat prescribing, computer use for morbidity data, computer use for audit, the receipt of prescribing data; the age of the senior partner; average age of GPs in the practice and percentage of GPs who are male. Some univariate differences were statistically significant, but since no attempt was made to control for covariates, we do not know if these differences would remain significant in a multivariate context.

In conclusion, a wide array of measures of physician practice style have been attempted, many of which are not statistically significant. Those that have been statistically significant most often include measures of physician age or experience (four out of eight studies), gender (two out of five) and specialty of physician (three out of six). The prescribing studies identified fundholding status and some quality indicators such as percentage of generic prescribing and using a practice formulary as significant determinants of prescribing expenditure.

It appears that current variables are not sensitive enough to detect the effect of practice style on utilisation. As Reid et al. (1999: 98) state, regarding referral behaviour:

> *"Future studies should explore whether the remaining variation can be explained by psychological and sociological factors relating to the thinking and behaviour of individual practitioners and the interaction between physician and patient. Reasons previously suggested for variation in referral behaviour include the ability to live with uncertainty, ability to manage patient pressure, relationships with local consultants and previous complaints from patients'.*

2.2.2.5 Summary

Demographic characteristics always play a role in risk-adjustment studies. Age is always included, while its non-linear relationship with utilisation can be accounted for using a number of specifications. The sign on gender is usually but not exclusively positive, nor is it always significant. Other variables included occasionally but not usually in either risk-adjustment or other utilisation studies include marital status, ethnicity and family composition. The effect of marital status is ambiguous, while the justification for including family composition varies from study to study, as does its effect.

Health status can either be directly measured or inferred from health service utilisation. Individual-level risk-adjustment studies have regularly applied disability as a measure of health. Clinically measured variables and self-reported health status have been applied less frequently and with less success. Small area-level risk-adjustment studies have applied low birthweight and responses to Census questions as measures of health status, the latter type with more success than the former. Death or SMR can be a useful risk-adjuster in individual-level studies, especially if a mixture of prospective and retrospective payment is acceptable. It is often used in small-area risk-adjustment studies.

Diagnostic data can be based on inpatient or outpatient data or health status can be inferred from prescriptions data. All three are good predictors of future health care expenditure. Problems with perverse incentives and mismeasurement of health status have been identified and steps taken to minimise their effects. In addition, previous expenditure has been used as a risk-adjuster and amounts to partial capitation, or a combination of prospective and retrospective reimbursement.

Socio-economic variables are included occasionally in risk-adjustment studies to modest effect, unlike in demand studies where price and income effects are the focus of the studies. Despite the importance of time prices for a service free at the point of delivery, such as the GMS, distance is rarely included in health care utilisation studies. Where it is included, it is negative as expected. Finally, while large variations in physician practice style have been identified, attempts to measure it have proved far less successful.

2.2.2.6 Implications of Previous Research

The Grossman model (1972) predicts that the demand for health care depends on the wage rate, the price of medical care, the stock of human capital and age. The importance of the physician is also highlighted in theoretical work. Meanwhile the empirical studies reviewed above contain variables that were categorised into five types – demography, socio-economic circumstances, health status, access to health care or characteristics of the health care market and finally physician characteristics. We draw from both the theoretical and empirical models of health care demand to formulate this study's empirical specification as outlined in Table 2.2.

Table 2.2

List of Covariates

Type of Variable	Measure
Demographic	Age Gender Marital status
Socio-economic	Lone parenthood Disability Number of people on each medical card
Health status	Chronic Illness
Access	Distance to GP Rural residence Registered with GP in receipt of rural practice allowance
Physician Characteristics	GP age Nurse Secretary Panel size Proportion of prescribing expenditure that is 'specific', 'symptomatic' and 'often presumptive'

The demographic variables that we include are age, gender and marital status. The theoretical model predicts that age is positively associated with health care utilisation because it measures the rate of health status depreciation. Empirical studies also find that age is positive, although in a non-linear manner. The eventual specification of the age variable is discussed in more detail in Chapter 4 as part of exploratory data analysis.

Although it was not directly implied by the theoretical model, Grossman (1972) included gender in his empirical model. The empirical literature produces no consensus on the sign or the significance of the gender variable. It was negative in one risk-adjustment study and one demand study, positive in three other demand studies and insignificant in a number of other studies. Therefore, we cannot form a strong view on the likely effect of this variable, so it can never be assigned as 'counter-intuitive'.

The inclusion of marital status is not implied from the theoretical model, but has been included in a number of empirical studies. In two of those reviewed above, it is insignificant (including one risk-adjustment study) and in one being single in negatively associated with visitation rates. Therefore, we test the hypothesis that

being married is positively associated with utilisation, although we concede that its sign is ambiguous.

The role of socio-economic status on health care utilisation is established theoretically and empirically. Grossman (1972) stresses the role of human capital (measured as education) in the efficient production of health status. There is also a large body of literature stressing the role of poverty and socio-economic status (including low educational achievement), in producing ill-health, reviewed comprehensively in Department of Health (1999), such that the schooling effect could be measuring socio-economic status rather than human capital accumulation. Thus, there are three propositions for the causal relationship between education and health. First is that causality runs from education to health. Education provides one with information on how best to produce health, such as through abstinence from smoking. Moreover, it can affect one's tastes and preferences for health. Second, causality can run from health to education, with healthier people better able to succeed at school. Third, there may be no causal relationship, with educational attainment representing socio-economic position, which then affects the production of health. Grossman (2000) reports that the weight of evidence favours a causal relationship from schooling to health. One remaining empirical issue, reviewed by Grossman (2000) is that education may actually be a measure of time preference, the suggestion being that those who invest in more education have a lower discount rate, placing a greater weight on their future health status and therefore being more likely to engage in healthy behaviour, such as abstinence from smoking. This hypothesis still needs rigorous empirical testing according to Grossman (2000).

Thus, socio-economic status should be included in our study because it is so well established empirically that people in more deprived circumstances have worse health and higher health care utilisation. In the terms of Grossman (1972) the inclusion of socio-economic position can be justified in so far as it measures stock of human capital, with deprived people having a lower stock of human capital and therefore a relatively inefficient health production function, implying a greater demand for health care.

Although we neither have a measure of education or social position in our dataset, we have data on lone parent family payments and number of people on the medical card. We expect that coming from households in receipt of lone parent family payments is positively associated with health care utilisation, since lone parenthood is a well established measure of poverty (Callan et al., 1996; Department of Health, 1999).

The number of people on the medical card is a measure of family size. Family size is positively associated with poverty, which is the motivation for including this variable. For instance, Deb and Trivedi (2002) suggested that their family size variable indicated unobserved financial distress. We expect it to be positively associated with health care utilisation. However, it has a second potential interpretation. Larger families could be associated with higher childcare costs implying higher opportunity costs of accessing health care. Therefore, the variable could actually be negatively related to utilisation. Meanwhile, Hornbrook and Goodman (1995) produce a similar family composition variable, which they interpret as a representation of family decision making. It is unclear what interpretation is to be placed on their results based on these grounds for inclusion. Therefore the original motivation for including this variable suggests a positive sign, but given the second potential interpretation, its sign is unclear.

In the terminology of the Grossman model, having a chronic illness is akin to having a higher health status depreciation rate. We include measures of chronic illness based on chronic disease score methodology and expect them to be positively related to utilisation. As a second measure of health status, we include a dummy variable indicating individuals who come from a household in receipt of disability payments. This could be treated as a direct measure of health status, as per those risk-adjustment studies reviewed above, except that it does not indicate the person receiving the disability payments. Since the receipt of disability payments is also an indicator of poverty, this variable is a mix of poverty-related and health status-related determinants of utilisation. Either way we expect it to be positively signed.

The Grossman model predicts that higher medical prices cause lower utilisation. Although GMS services are free at the point of delivery, time prices are expected to

be important. We include distance to the GP, rural residence and, in the case of the model that includes supply-side variables, rural practice allowance, as measures of access to health care. We expect each to cause reductions in utilisation.

Phelps (2000) and McGuire (2000) among others stress the importance of variations in physician practice style, including physician effort, in determining health care utilisation. We found that although the evidence of the importance of the role of the physician is strong, there are few variables that appear to measure that effect well. Physician age is often significant, although its sign varies by study. We also include the presence of a nurse, a secretary and panel size as measures of practice quality, similar to Gravelle et al. (2002) and Whynes et al. (1997). The effect of practice quality on prescribing expenditure is unclear from the literature. They are included partly to detect and control for systematic variations in quality across GPs and partly to explore their effect on prescribing expenditure.

Given the poor performance of many measures of practice style in explaining variation in health care utilisation, we have derived a number of new indicators of prescribing style for this study. We wish to measure the physician's relative aversion or partiality for prescribing. According to McGavock's (1988) teleological classification of drug types reflecting perceived use in general practice, pharmacological treatment can be 'specific', 'symptomatic' or 'often presumptive'. Specific prescribing refers to prescribing where *"there is always an accurate diagnosis, often confirmed by laboratory test or other investigations and often by a specialist ... [and] the drug must be known to intervene in a specific and well-understood manner to alter the pathophysiology in the patient's favour"* (McGavock, 1988: 192). Symptomatic prescribing refers to prescribing where *"drugs relieve symptoms with no effect or slight effect on the disease process"* (McGavock, 1988: 193). Often presumptive prescribing refers to drugs which intervene in a specific and well-understood manner (as with specific prescribing) and which are ideally prescribed after a firm diagnosis has been reached, but which are often prescribed *"on the basis of a presumptive diagnosis on therapeutic 'trial'"* (McGavock, 1988: 193). The particular drugs that are included in each category are outlined in the data chapter below. We hypothesise that the proportion of total prescribing that is 'specific' is negatively related to an individual's prescribing expenditure, as it

suggests that a GP is more likely to initiate a prescription programme only when a positive diagnosis is established, that is, the GP who is prescribing averse. We expect that the proportion of total prescribing that is 'symptomatic' or 'often presumptive' are positively associated with an individual's prescribing expenditure, because both are indicators of a GP who is more partial to prescribing.

The concept of using the GPs prescribing behaviour as a measure of their prescribing style has been attempted before. Previous measures include percentage of generic prescribing (Whynes et al., 1997; Morton-Jones and Pringle, 1993; Reid, et al., 1999) and ratio of corticosteroids to bronchodilators (Reid et al., 1999), which have had mixed effects. The measures we propose are based on a much higher proportion of a GP's prescribing and as such may be better measures of prescribing style.

Finally, we include health board of residence as a control variable for health board effects.

2.2.3 Empirical Models of Utilisation: Econometric Considerations

The remaining element of equation (2.2) that needs to be discussed is its econometric specification. Health utilisation data are characterised by nonnegativity, a large proportion of the total sample and population that are zeros and a positively skewed distribution of the non-zero data, similar in many ways to earnings data. There is a large literature on econometric estimators that produce consistent and efficient estimates, given these awkward features of the distribution. We deal with each of these features, and candidate remedies, in turn. In addition, the dataset used in this study can be considered hierarchical, with groups of individuals forming GP panels, as is clear from the subscripts i and j in (2.2). We discuss how to model this feature of the dataset. Finally, as with any cross-sectional study, we need to consider potential endogeneity. We consider potentially endogenous covariates and review approaches to dealing with them.

2.2.3.1 Non-negativity and High Frequency of Zeros

The two-part model or left-censored models, including the generalised Tobit or Heckit, are attempts to deal with the first two features of most health utilisation datasets – non-negativity and a spike at zero.

The two-part model's approach is to identify the determinants of users of the service first, and in the second part to identify the determinants of expenditure by the users. The first part employs a dichotomous response variable and is usually modelled as either a logit or a probit (Diehr et al., 1999:135). The probit specification is as follows:

$$\text{Prob}[y = 1] = \Phi(\alpha x), \tag{2.21}$$

where y is one if the individual has positive prescribing expenditure and zero if the individual has zero prescribing expenditure, x is a row vector of covariates, α is the set of parameters to be estimated and $\Phi(.)$ is the standard normal distribution. Modelling the second part is discussed below but can be expressed in general terms as:

$$y/\, y>0 = x\beta + e, \tag{2.22}$$

where β is a set of parameters to be estimated and e is the error term

The expected expenditure by any individual, therefore, is the probability that they will use the service multiplied by their expected expenditure, given that they are users, i.e.

$$E(y \mid x) = \hat{P}\, E(y \mid y > 0, x), \tag{2.23}$$

where $\hat{P}$ is the probability of getting a prescription from the first part of the model. Although the two-part model is justified chiefly for statistical reasons, it also can be viewed as the product of an attractive behavioural model. The first equation can be used to model the decision to attend to GP, while the second equation can be used to model the decision on the extent of health service utilisation for the proportion of the population who are users. Therefore the first part of the model focuses principally on the decision-making of the patient, while in the second part of the model the decision-making focus shifts more towards the GP, who generally decides how much medical treatment is appropriate for the patient.

The specification bears a close relation with left-censored models such as the sample selection model (Heckman, 1979). The behavioural analogue of the two-part model implies a sequential decision-making process. The patient decides to attend the physician and the physician or physician and patient together decide how much health service utilisation is required. Meanwhile, the generalised Tobit model is characterised by the decision to consume and the amount of consumption being made jointly. Actual decision making processes in health care fall somewhere between these poles. For instance, the decision to attend is affected by the severity of illness and hence expected utilisation. As such, the decision to attend and the decision on the extent of utilisation are at least to some extent a joint decision, rather than purely sequential.

There has been considerable debate as to the relative merits of the two-part model versus a generalised Tobit or sample selection model in health econometrics. Jones (2000: 289) points out that the sample selection model performs poorly if the parameter estimates from the selection equation and the explanatory variables from the demand equation are highly correlated, which can occur if there are few exclusion restrictions; the proportion of censored observations is large; the range of the explanatory variables in the selection equation is small or there is a large residual variance in the selection equation. According to Jones (2000:289), collinearity problems are likely to arise in health datasets, so the sample selection model should only be applied once collinearity tests have proved satisfactory. Although the sample selection model has been used in applied health economic research (Zimmerman Murphy, 1987; Hunt-McCool et al., 1994), the two-part model has been used to a much greater degree (Pohlmeier and Ulrich, 1994; Street et al., 1999; Grootendorst, 1995). As such, we apply the two-part model in this study.

Another alternative to the two-part model approach is the use of finite mixture models (Deb and Trivedi, 1997, 2002). These test for a mixture of conditional distributions in the dataset, with between one and four distributions being common in many microeconometric datasets. Since we have no prior expectation of the number of distributions in the dataset and given the enormous computational time required to apply them, we explore finite mixture models further in Chapter 4 and decide at that point on their use as estimators of (2.2).

2.2.3.2 *Positively Skewed Data*

The most common approach to dealing with a positively skewed response variable is to transform the response variable and apply the following model:

$$\ln(y) = x\delta + \varepsilon, \tag{2.24}$$

where x are regressors including an intercept, δ are parameter estimates and ε is a residual. If ε is normally distributed $N(0,\sigma_\varepsilon^2)$, then $E(y \mid x) = \exp(x\delta + 0.5\sigma_\varepsilon^2)$. If ε is not normally distributed but is i.i.d., or if *exp(ε)* has constant mean and variance, then $E(y \mid x) = s\exp(x\delta)$, where the smearing estimate $s = E(\exp(\varepsilon))$. Duan (1983) proposed the mean of the exponentiated log-scale errors as a non-parametric smearing estimator, that is, $s = \sum^{n_+}[\exp(\varepsilon)/n_+]$, where n_+ represents the sub-sample of positive values in the dataset. This is the approach taken by the Rand Health Insurance study (Duan et al., 1983; Manning et al., 1987).

However, if ε is heteroscedastic in x, then $E(y \mid x) = f(x)\exp(x\delta)$ or in the log normal case $\ln E(y \mid x) = x\delta + 0.5\sigma_\varepsilon^2(x)$. Therefore, the smear estimator s will be a biased estimator of *E(y|x)* in a way that will depend on x. This problem can be overcome by including an estimate of the variance function *e(exp(ε)|x)*, or if the error term is log normal $\upsilon(\varepsilon \mid x)$. One such approach is described below.

Despite problems such as heteroscedasticity and non-normality, Manning and Mullahy (2001) outline five different estimation strategies and assess their consistency and precision in estimating the retransformed response variable, for a variety of simulated data generating processes. The first estimator is an OLS regression of *ln(y)* on x and an intercept using Duan's (1983) homoscedastic smearing factor. The second is an OLS regression of *ln(y)* on x and an intercept using a heteroscedastic smearing factor, specified as follows:

$$\upsilon = E(\varepsilon)^2 = \delta_0 + \delta_1 x + \delta_1 x^2. \tag{2.25}$$

In other words, the heteroscedastic smearing factor uses the predicted value of the errors, while the homoscedastic smearing factor uses the mean of the errors.

The other three estimators are generalised linear models (GLM). The advantage of the GLM is that it can allow for non-constant variance while maintaining the original scale of the data. As described in Blough et al. (1999) there are three parts to the GLM. First, there is the linear predictor η, similar to OLS:

$$\eta = x\gamma \,, \tag{2.26}$$

where γ is a set of parameter estimates. Second, there is a monotonic differentiable link function *g()* which maps the linear predictor onto the expectation of the response variable:

$$g(\mu) = x\gamma \,, \tag{2.27}$$

where $\mu = E(y)$. Third, there is a variance function *V()* describing how the variance depends on the mean as follows:

$$Var(y) = \sigma^2 = \kappa V(\mu) \,, \tag{2.28}$$

where κ is a constant called the dispersion parameter. An array of link functions and variance functions can be specified. The logarithmic link is popular for health expenditures (Blough et al., 1999; Madden et al., 2000; Manning and Mullahy, 2001). Therefore, an individual's predicted expenditure can be computed simply as:

$$E(y) = \hat{P}\exp(x\hat{\gamma}) \,, \tag{2.29}$$

where $\hat{P}$ is the probability of use from the first part of the two-part model.

Manning and Mullahy's (2001) final three estimators differ only in how the variance is specified. The first GLM has a variance that does not depend on *E(y\x)* or *x*, which is the non-linear least-squares estimator proposed by Mullahy (1998) to deal with positively skewed data. The second GLM estimator has a variance that is proportional to *E(y|x)*, which is a Poisson-like distribution with overdispersion. The third GLM has a standard deviation proportional to *E(y|x)*, which is a gamma-like assumption similar to the model proposed by Blough et al., (1999). As such, Manning and Mullahy (2001) test the three estimators previously proposed to overcome problems with positively skewed non-negative data (Manning, 1998; Mullahy, 1998; Blough et al., 1999), as well as the 'naïve case' - the first estimator above - and a GLM with an overdispersed Poisson-like set of assumptions.

Manning and Mullahy (2001) examine each of the above estimators in treating skewed data; heavy-tailed data; monotonically decreasing or skewed bell-shaped probability density functions; heteroscedastic data and data that are modelled exponentially but have an additive error term on the untransformed scale. They found that the GLM models were best at overcoming retransformation bias. However they often did not have adequate precision, that is, depending on the data generating process, GLM models could be inefficient. Therefore, they produce an algorithm for model selection that starts with a GLM estimator and chooses between the various competing estimators depending on the kurtosis and skewness of the log-scale and untransformed errors.

Despite the features of non-negativity, a spike at zero and positively-skewed positive values in the datasets used in most – perhaps all – risk-adjustment studies, by far the most common estimation technique is a simple OLS model (van de Ven and Ellis, 2000: 788)[5]. Van de Ven and Ellis (2000:788) suggest three reasons for the preference for simple OLS. First, heteroscedasticity on the log-scale is a feature of most risk-adjustment datasets, meaning that retransformation using techniques such as a Duan smear estimator would produce biased estimates. Second, risk-adjustment datasets are routinely extremely large. They refer to a note in Mullahy (1998) stating that when sample sizes are large, linear regression may be an acceptable alternative estimation strategy to transformation of health utilisation data with the features described above (although Mullahy (1998) does state that linear regression may be an acceptable alternative, we found no reference associating it with sample size). Third, the simple linear model remains close to the calculation of average expected expenditure per risk-group, which is the approach for premium rating used by health insurance companies and health plans.

Referring to the first point, recent advances in dealing with the retransformation problem in the presence of heteroscedasticity as outlined above make alternatives to the simple OLS model more attractive. Referring to both the first and second points,

[5] Exceptions include Dunn (1998) and Madden (2000).

these are essentially empirical considerations. According to Diehr et al. (1999:135) *"[i]f the data set is very large, OLS regression on the untransformed data (including the zeros) will provide unbiased estimates of the regression parameters. The standard errors of regression coefficients may be too large...however, in the large datasets often available, significant effects are usually so strong that doubling or tripling the standard error would have little effect on the conclusions"*. Large datasets undoubtedly overcome the potential inefficiency of simple OLS. However, in a risk-adjustment study examining employed and disabled populations in the US, Madden (2000) compared simple OLS to a two-part GLM model with a log link function and a gamma variance function and found that the GLM model performed much better. As Mullahy (1998:279) puts it, *"[t]he question of whether 'truth' is really a one-part or a two-part model should be confronted squarely in applications"*. In the empirical chapter below, we assess a number of these estimators against a set of model selection criteria.

2.2.3.3 Hierarchical Modelling

The dataset used in this study can be viewed as hierarchical. In some cases individuals are grouped into households; individuals, and again in some cases entire households, attend one particular GP; GPs in most cases belong to one health board area, although GPs at health board borders may have many patients from outside the health board of residence of the GP. Therefore, at least a two-tier hierarchy can be identified, as identified by the inclusion of subscripts i and j referring to individual i and GP j in (2.2).

Aside from the inclusion of indicators of health plan, hierarchical models have not been applied in risk-adjustment studies that use individual-level data. First, the interest in these studies is in consistent and efficient estimation of the individual-level explanatory variables. The economic significance of higher level variances is of minor importance. Second, complex hierarchical designs are computationally time-consuming. With the large datasets used in risk-adjustment studies such as this one, complex hierarchical designs are never employed. Thus, in order to relax the assumption of independence of individuals attending one GP, we employ cluster-robust standard errors. This is a similar formula to the more familiar Huber-White heteroscedasticity robust estimator. The Huber-White estimator ($\hat{H}$) is:

$$\hat{H} = \hat{V}(\sum_{i=1}^{N} u_i' u_i)\hat{V}\,, \tag{2.30}$$

where $\hat{V} = (-\partial^2 \ln L / \partial \beta^2)^{-1}$, u_i is the contribution from the i^{th} observation to the scores $\partial \ln L / \partial \beta$, $\ln L$ is the log likelihood and β is the vector of regression coefficients. Meanwhile, if the observations are not independent but can be grouped into J groups G_1, G_2, ... G_J that are independent, then the cluster robust estimator can be derived as follows:

$$\hat{H} = \hat{V}(\sum_{j=1}^{J} u_j' u_j)\hat{V}\,, \tag{2.31}$$

where u_j is the contribution from the j^{th} observation to the scores $\partial \ln L / \partial \beta$. Essentially, the cluster robust estimator inflates standard errors to reflect the fact that observations are not completely random.

2.2.3.4 Endogeneity and Unobserved Heterogeneity

"In a cross section, predetermined variables can rarely legitimately be treated as exogenous" (Deaton, 1997: 99). If one of the x variables in $y=x\beta+u$ is determined by factors that include y then u will be correlated with one or more of the x's and OLS estimates will be biased and inconsistent. Most small area analysis of health care utilisation in the UK, including Carr-Hill et al. (1994) and Rice et al. (2000), consider the interdependence of supply of NHS services with need for NHS services – in areas of high need, the supply of health services may be greater (assuming a direct rather than an inverse care law), so that high utilisation in areas of high need may simply be due to lower access costs in areas of high need rather than high need leading to high utilisation. In these studies, two stage least squares is used to estimate the independent effect of needs variables on utilisation, controlling for supply. This is rarely a cause for concern with micro data, where one individual cannot affect the supply of services (Schokkaert and van de Voorde, 2000) although endogeneity in essentially an empirical issue.

Acton (1975) views endogeneity working the other way around, whereby individuals choose to live close to health services because they are high users (rather than health services locating in areas where there is high use, which was the concern of the UK research). It is the one exception in individual-level studies where an individual's

distance to ambulatory care was treated as an endogenous variable, which led to a two-stage least squares analysis. As reviewed by Van de Ven and Ellis (2000), risk-adjustment studies do not consider endogeneity of distance to services, either due to high utilisation individuals choosing to live near health services, or health services locating in areas where there are a lot of high need individuals.

Error in the measurement of chronic illness using previous utilisation of health care may be endogenous. In order to be indicated as suffering from a chronic illness, you need to have access to health services and your GP has to identify the chronic illness. Thus you may have a high level of prescribing expenditure and be indicated as having a chronic illness simply because of the prescribing style of your GP. The issue of endogeneity of measurement error, and ways to treat it, is discussed in detail in Chapter 3.

2.2.4 Conclusion and Modelling Strategy

The introduction to this chapter outlined two gaps in the risk-adjustment literature, because of the critique of the standard risk-adjustment model, especially by Schokkaert and van de Voorde (2000). These were a theoretical basis for explaining the relationship between utilisation and a set of covariates and a comprehensive review of the variables that should be included in a properly specified utilisation function. Section 2.2.1 generated predictions from theoretical models relating to individual behaviour and physician behaviour. These were supplemented with empirical findings on expected effects from empirical literature in section 2.2.2, such that we generated an expected sign for each covariate in section 2.2.2.6, or identified covariates where expected effects are not clear. Section 2.2.2 also outlined the difficulty in measuring adequately concepts such as health and physician's practice style. Consequently, we proposed new measures of physician's prescribing style in section 2.2.2.6. Section 2.2.3 outlined a number of features of health utilisation data that have led researchers to using alternative estimators to OLS. These considerations inform our modelling strategy.

A number of models are proposed and tested. As described in Chapter 1, up to the beginning of this study, the IDTS was based largely on the national age-related average prescribing expenditure in each of six age categories, the so-called NARA.

The discussion in Section 2.2.2 concluded that additional risk-adjustment variables to be considered include demographic and socio-economic variables, health status variables, access to health services and supply-side variables. However, we found in section 2.2.2 that using previous utilisation to measure health status can lead to a perverse incentive structure arising, while there is no guarantee that they are unbiased measures of health. Although we reviewed steps taken to minimise or eliminate these two problems, the steps taken may not be sufficient for some policy makers. We also reviewed literature that included supply-side variables in utilisation functions. We found that many of these variables were poor measures of physician practice style and proposed a number of new ones, based on McGavock (1988). Again, some policy-makers may be uncomfortable with the inclusion of these new variables, whose validity is not as well established as other variables in the model. Therefore, in addition to the NARA, three models are proposed. First, we estimate one model containing age, gender and a vector of socio-economic variables. We refer to this as the Demographic model, since that is the term applied to a similar model in Dutch risk-adjustment (Lamers and van Vliet, 2001). The second one consists of the Demographic model as well as indicators of chronic illness. We refer to this as the Chronic Illness model. The third one consists of the Chronic Illness model as well as supply-side indicators. We refer to this as the Supply model. These are estimated using OLS, similar to the majority of risk-adjustment models, which use very large datasets to overcome the unhelpful features of health utilisation data described above. Since individuals form GP panels, we relax the assumption of independence of observations within a GP panel by applying cluster-robust standard errors.

However, despite the popularity of OLS, a number of alternative estimators were proposed in section 2.2.3, which concluded that the choice of model is an empirical issue. Therefore, an additional set of models is also estimated. First, there is the set of two-part models proposed by Manning and Mullahy (2002). We test the two-part model with the response variable of the second part transformed to natural logarithms. We then use either the Duan homoscedastic smearing estimate (Duan et al., 1983) or the Manning heteroscedastic smearing estimate (Manning, 1998) to retransform results into natural units for budget setting. Of the three generalised linear models examined by Manning and Mullahy (2001), the one with the gamma

variance function appears to be the best fit. It has also been used by Blough et al. (1999), Madden et al. (2000) and Deb and Burgess (2002). We also test it here.

Second, there is the sample-selection (or Heckit) model. However, owing to the problems outlined above, it is rarely now used in health utilisation modelling, so we choose not to apply it here. Third, there is the set of finite mixture models suggested by Deb and Trivedi (1997, 2002). Since it is unclear which, if any, of these finite mixture models performs best, and given the enormous computational time to apply them, we examine them further in the exploratory data analysis in Chapter 4. Therefore, we estimate at least seven models – the NARA, the Demographic model, the Chronic Illness model, the Supply model, the two-part logarithmic model with homoscedastic retransformation, the two-part logarithmic model with heteroscedastic retransformation and finally the two-part GLM. In addition, depending on the results of the exploration of finite mixture models in Chapter 4, additional models may also be estimated.

2.3 FROM UTILISATION FUNCTIONS TO BUDGET SETTING

The last section described a number of competing models of health care utilisation that have been applied empirically. The two principal criteria used for model selection are theoretical consistency and goodness of fit. Theoretical predictions were outlined in section 2.2.2.6. Measures of goodness of fit are discussed in Chapter 5. In addition, once we shift focus from estimating a consistent and efficient utilisation function to producing a formula for budget setting, a number of additional criteria arise. At the level of the individual, these include predictive ability and distributive implications. Measures of both are discussed in Chapter 6. At the level of the GP, in addition to the effect of aggregation on goodness of fit, predictive ability and distribution, we examine risk exposure of the budget holder. Measuring risk exposure is discussed in Chapter 7.

While measurement issues are postponed to the empirical chapters below, this section provides an overview of the literature on distribution and on risk exposure, discussing the rationale for including both as criteria for selecting a budget setting formula.

2.3.1 Distribution

We wish to ensure that the distribution of health care resources is equitable. Thus, we first define equity; then we examine the application of the definition to risk-adjustment.

There are two components to equity, horizontal equity and vertical equity. Horizontal equity is the equal treatment of equals while vertical equity is the unequal treatment of unequals. Equity is one of the four guiding principles of the Irish health strategy, which states that *"equity will be central to developing policies (i) to reduce the difference in health status currently running across the social spectrum in Ireland; and (ii) to ensure equitable access to services based on need"* (Department of Health and Children, 2001:18).

The first part of the statement, concerning how socio-economic differences in health might be addressed, reflects considerations of vertical equity. The second part of the statement can be interpreted as reflecting horizontal equity considerations. Hence,

the operational definition of equity used in the Irish health care setting focuses both on vertical and horizontal equity.

While the concept of horizontal equity is reasonably intuitive, vertical equity is less so. Essentially, incorporating vertical equity considerations into health service objectives means that we wish to make the distribution of health care resources more progressive, that is the amount of health care resources per needs-adjusted head of population should increase with increasing need. For instance, suppose that when a hypothetical index of need is 100, health care expenditure is IR£1000, meaning that the average expenditure per needs score is IR£10. When the need index is 200, expenditure is IR£4000, meaning that average expenditure per needs score is IR£20. This is a progressive regime. Progressivity essentially means a departure from proportionality. Treating unequals unequally involves first measuring the extent of the inequality and second choosing the rate at which we positively discriminate in favour of the disadvantaged. The latter clearly involves value judgements, although it can be illustrated that there are value judgements in all inequality measurement.

The vast majority of weighted capitation models focus purely on horizontal equity. Indeed Sutton and Lock (2000) showed that the weighted capitation methodologies employed in the UK took no account of vertical equity in the design of the models. However, one of the principal drawbacks with the 'empirical approach' to weighted capitation is that it is assumed that the level of unmet need does not differ systematically across the population. As Carr-Hill et al. (1994, p.138) acknowledge:

> *"This entire study was predicated on the assumption that utilisation of NHS inpatient resources is a good predictor of health care need. For many reasons, this assumption may be suspect. Some groups may be systematically excluded, others may 'capture' more NHS resources that their clinical need justifies. There is a clear need for research to establish whether utilisation is a legitimate predictor of need"*

There is a considerable body of evidence to suggest that some groups face significant barriers to access to health care. Reasons for differential access include the 'inverse care law' (Tudor Hart, 1971), a stylised fact that proposes that the availability of health care is in inverse proportion to its need. In Ireland, Nolan and Russell (2001), in an analysis of the distribution of medical card benefits (meaning free health care) by income decile, found that while the majority of medical card benefits accrue to the

lowest three income deciles, their distribution is not even within this group, with the third lowest income decile getting 23.2% of total medical card benefits and the bottom income decile getting 17.7% of total benefits. Given our expectation that lower income is associated with lower health and therefore greater need for health care, we expect that the lowest income decile would have higher health care expenditure than the third from lowest. The data are not standardised for need, however, so any analysis must bear this in mind.

Meuller (1984) found that the rate of use of dental services by poor children with Medicaid in the US was well below the norms for children with private insurance. In addition, lower class patients tended to receive less expensive treatments than others (Hazelkorn, 1985). Simply equalising the financial cost of health services does not equalise utilisation for equal levels of need for each social group.

There is a large body of evidence relating to differential access to specialist services. For instance, in Scotland, rates of coronary artery bypass grafting and angioplasties within two years of an acute myocardial infarction were significantly greater among the most affluent socioeconomic group, as measured by the Carstairs and Morris deprivation category, when compared to the rest of the population, despite lower incidence of coronary heart disease and greater access to private health care (McLaren and Bain, 1998). In Sweden, Diderichsen et al. (1997) found that non-Nordic immigrants used mental health services less than the Swedish, despite evidence of having higher levels of mental morbidity. Chaturvedi et al. (1995) examined GP consultation rates and hospital procedure rates for selected conditions by socio-economic group for 1981/82. They found that for some conditions there were higher rates of GP consultations by lower social classes, but no difference in operation rates with the rest of the population. The relationship was the opposite in the case of hip replacement, with procedures being carried out on the relatively affluent more than the relatively deprived, despite the latter consulting their GPs more on the condition. Payne et al. (1997) examined the relationship between need, supply and demand on rates of revascularisation for the treatment of angina in Sheffield. Need was assessed using the Rose angina questionnaire, sent to over 16,000 residents of Sheffield, none of whom were more than 20km from a major cardiology centre. Moreover, specialist investigation and treatment centres were

located in some of the most deprived wards of Sheffield, as measured by Townsend score. However, deprived areas had approximately half the number of revascularisations per head of population with angina compared with affluent wards. Revascularisations per premature death from coronary heart disease and myocardial infarction were also compared to Townsend score, with a similar inverse relationship prevailing. These studies suggest that deprived populations receive less specialist attention than more affluent populations for equal levels of need.

While the evidence supporting the unmet need hypothesis is convincing, there is very little theoretical literature in health economics describing why unmet need occurs at the levels described above. Some hypotheses for this are inequities in the distribution of GPs (Sinclair, 1999; Gravelle and Sutton, 1998), and inequities in the distribution of resources available to GPs, including prescribing budgets, access to secondary care and access to other primary care services, such as public health nursing (Gravelle and Sutton, 1998). It remains more a stylised fact than a theoretical prediction. In sociology, Grembowski's (1989) Social Exchange Theory suggests that behavioural, cultural, provider and environmental factors appear to be more important than the need for care in determining whether under-served groups visit the health services. A clear economic theory for unmet need would help in determining how best to tackle it from an economic perspective.

Based on the empirical evidence indicating unmet need, there is a case for examining it before proceeding with utilisation based models of risk-adjustment. If deprived populations under-utilise services for any given level of need, utilisation data will underestimate the need of deprived populations, and GPs with large proportions of deprived populations will be funded less than the true needs of their populations require. However, Rice and Smith (1999) argue against the use of an unmet need adjustment, since there is no guarantee that directing resources towards providers who have greater populations of disadvantaged people will result in greater access to health care for disadvantaged people. This is, however, simply an argument against the use of unmet need adjustments on their own. If linked to appropriate monitoring mechanisms including the measurement of GP's progressivity, it is likely that they could improve access for disadvantaged groups.

Aside from the normative approaches (which measure need directly and therefore do not suffer from the unmet need problem), we only know of two current risk-adjustment exercises that account for unmet need. One is in New Zealand, where there is evidence that the Maori population under-utilise health care services. Therefore, resources are allocated on the basis of how much the Maori population would be expected to consume if it consumed at the same rate as the non-Maori New Zealand population for any given level of need, as measured by under-65 SMR (Rice and Smith, 1999).

The Scottish Executive (1999) also propose an unmet need adjustment to its weighted capitation formula. Two approaches to limiting unmet need effects are outlined. The first is to find that area where access to services is maximised for all groups and apply that area's results nationally. The second is to identify the relationship between need and use for socio-economic groups who do not suffer relatively from poor access, and apply these results to all groups (similar to the treatment of Maoris in New Zealand).

Sutton and Lock (2000) measured the extent to which health care use rose with increasing need for Scottish Health Boards using a Kakwani index of progressivity. The slope parameters of the most progressive Board were then applied nationally, which then formed the basis of hypothetical budgetary re-allocations.

A second approach suggested by the Scottish Executive (1999) is similar to the Maori adjustment outlined above. The relationship between need and use for the members of the population who are not socially excluded is estimated and this is used to generate the needs index for the whole population. Lock and Sutton (2000) found that the positive relationship between need and use tailed off with greater levels of deprivation, suggesting that unmet need was in evidence for socially excluded groups.

Both these approaches to adjusting utilisation-based weighted capitation formulae for unmet need are considered again in Chapter 6 and in Appendix 6.1.

2.3.2 Risk Exposure

Many risk-adjustment exercises explain only about 20% of the total variation in health care expenditure between individuals (Newhouse et al., 1989). On the other hand, risk-adjustment of groups is much more predictable. If high-risk patients are removed, up to 99% of health care costs can be predicted successfully (Robinson et al., 1991), while Newhouse et al. (1989) were able to predict 60% of future costs of groups. Nevertheless, the decision on who assumes the risk for absorbing the residual variation is critical.

In health care systems where health plans compete, such as the Irish health insurance market, the residual variation means that plans have an incentive for risk selection, that is, select against individuals who can be identified as being bad risks, even given their risk adjusted capitation payment. BUPA (Ireland)'s initial plan to offer cash bonuses to members aged less than 50 could be seen as an attempt to risk select. For GMS patients, health boards must offer treatment to all eligible residents of their area, so risk selection is almost impossible. Meanwhile, it is unlikely that GPs in Ireland undertake sophisticated risk-selection strategies, as the incentives are not very strong, while there is a risk of damaging collegiate relations if a GP is found to be risk-selecting. However, risk selection can be subtler, such as shifting the cost of treating a high cost patient from primary care to secondary care or vice versa. This is *de facto* refusal to treat, or *de facto* risk-selection. Thus, we need to consider risk exposure, not purely because it may be 'unfair' to expose the budget holder to excessive risk, but because it may lead to a set of unintended outcomes, or in the extreme the collapse of all contracts due to risk selection.

Newhouse (1998) recommends partial capitation, or supply-side cost sharing. This reduces the profit to be earned from engaging in risk selection and given that risk selection is not costless, well designed partial capitation systems would eliminate the incentive for risk selection. Moreover, partial capitation removes the incentive on plans to stint in the provision of services because, under partial capitation, plans will receive additional funds with additional provision of services, whereas under full capitation no extra funds are received.

The rationale for partial capitation is akin to that for cost-sharing on the consumer side. The extent to which individuals will need to spend on health care in the future is so uncertain that risks are pooled using health insurance schemes (of which public finance is a type). Cost sharing is then introduced to reduce the problem of moral hazard, that is, insured parties engaging in riskier behaviour once they have taken out insurance. Health plans face a similar problem of unpredictable future health care costs. Newhouse's proposed remedy is to share this risk between the plan and the central funder, while minimising moral hazard of the plan (they are now insured against unpredictable future risk) by engaging in cost sharing.

Van Barneveld et al. (2001) describe the elements of supply-side cost sharing, which they refer to as risk sharing, in some detail. There are four parameters to be considered. First risk sharing can be prospective or retrospective. Prospective risk sharing is where individuals whose costs will be shared are identified at the start of the year. They can be patients with specific high cost conditions such as AIDS sufferers or the terminally ill, patients with previous high costs, or patients identified in advance by the GP. Van Barneveld et al. (2001) recommend that the prior identification of high risk patients by providers is likely to be a more efficient scheme than condition-specific risk sharing, as condition-specific risk sharing may lead to providers falsely recoding patients as suffering from a condition for which the risk is shared, similar to 'DRG-creep' in the hospital setting.

Retrospective risk sharing occurs when individuals whose risks will be shared become known during the year. With retrospective risk sharing, prospective financing would be based on the risk-adjustment model. The actual spend may then be unavoidably higher or lower than the predicted spend. A retrospective adjustment is introduced as a remedy. There is currently a 'risk-equalisation' fund in place amongst Ireland's two health insurers VHI and BUPA (Ireland) which retrospectively compensates one if they have accepted more high risk patients than expected. The same principle could be applied to GMS prescribing budgets. Retrospective risk sharing is a feature of the Dutch health care system.

The second parameter that van Barneveld et al. (2001) consider is the types of care for which risk is shared. For instance, catastrophic care might be shared. In the

IDTS in Ireland, certain high cost drugs are 'budget neutral' meaning that they are not included in the GP's budget. However, it is possible that the costs for these drugs could be shared between the GP and the GMS (Payments) Board. This introduces the third parameter, which is the extent of risk sharing. Van Barneveld et al. (2001) describe three common options applied here, namely proportional risk sharing, outlier risk sharing and normative risk sharing. Proportional risk sharing is where a proportion of all costs are assumed by the financier. In the case of 'budget neutral' drugs in the IDTS, that proportion is one. Outlier risk sharing is where risk sharing occurs above a certain threshold, also known as a stop-loss arrangement. This is a another feature of the IDTS, where 'high cost' patients are identified and the GMS (Payments) Board covers all costs above the specified 'high cost' threshold. This was set at IR£2075 in 2000. Normative risk sharing occurs when an individual's cost exceeds their predicted cost according to the capitation payment.

The fourth parameter is the price paid by the provider for risk sharing. In Ireland, the GMS (Payments) Board top-slices a certain amount of their budget to cover high cost patients and budget neutral drugs. The budget given to a GP is reduced by the amount top-sliced by the GMS (Payments) Board, multiplied by the GP's budget as a fraction of the total remaining budget. This can be treated as the price paid by the GP for insurance.

One final risk management scheme to which van Barneveld et al. (2001) do not refer is pooling of risks amongst GPs. Martin et al. (1998) highlight the risks of devolving acute hospital budgets in the UK to too small a population, where the effects of a few high cost patients cannot be moderated by a large number of low cost patients. Some GPs in Ireland with panels of less than 100 have indicative drug targets. There may be scope for developing co-operative arrangements between GPs so that their budgets are pooled, thereby reducing risk. This is explored further in Chapter 7.

Thus, the current IDTS makes a number of adjustments for risk exposure of the GP. First, some drugs are assigned as budget neutral. In van Barneveld et al.'s (2001) terminology these drugs are subject to proportional risk sharing with a proportion of one. Second, the scheme retrospectively applies outlier risk sharing, again with a proportion of one. Clearly a number of other options are available. High risk

patients could be identified prospectively. The drugs included in the budget neutral list could be reviewed. The proportion of expenditure that is shared for high cost and budget neutral drugs could be reconsidered or the extent of risk sharing could be changed in other ways. The amount top-sliced by the GMS (Payments) Board could be renegotiated. Some of these aspects of risk exposure and risk management are examined empirically in Chapter 7.

2.3.3 Conclusions

This section outlined the criteria used for evaluating models. These included theoretical consistency, goodness of fit, predictive ability, distributive implications and risk exposure. Measurement issues were postponed until the empirical chapters below, while it may not be obvious initially why final two criteria are included in model selection. Therefore, this section examined the role of equity in the distribution of outcomes and risk exposure in budget setting. We found that there were two elements to equity - horizontal and vertical. In the absence of systematic unmet need, a risk-adjustment formula based on utilisation may be equitable. In the presence of unmet need, adjustments may be required. The candidate adjustments were reviewed. Finally since an individual's health care utilisation is subject to considerable random variation, budget holders can be exposed to significant risk. This may damage the budget scheme in two ways. First, it may simply be unfair and, secondly, it may lead to budget holders engaging in risk selection. Five considerations were reviewed in the management of risk. These were retrospective versus prospective risk sharing; type of care for which risk is shared; proportion of total expenditure for which risk is shared; price paid for risk sharing by the budget holder and sharing of risk between budget holders through budget consolidation.

2.4 CONCLUSIONS

This chapter examined the requirements for estimating the empirical approach to risk-adjustment. We found two gaps in the risk-adjustment literature as a result of the critique of the standard model by Schokkaert and van de Voorde (2000) and Carr-Hill et al. (1994). First, an explanation of the expected effects of each covariate, from first principles, had not been undertaken. Second, the absence of a comprehensive review of the variables to include in a utilisation function in the risk-adjustment literature was evident. Sections 2.2.1 and 2.2.2 attempted to fill these gaps. Theoretical predictions from the Grossman model and from studies on physician behaviour were supplemented with empirical studies leading to a set of hypothesised effects for each variable included in the empirical chapters below. The examination of the variables included in previous empirical utilisation functions categorised variables into five types – demographic, socio-economic, health, access to health services and physician practice style. We concluded by outlining the variables used to measure each of these concepts in this study, and the expected effect of each one.

Econometric considerations were outlined in section 2.2.3, drawing both on previous risk-adjustment studies and other studies of health care utilisation. We found that the choice between the one-part and two-part models was essentially an empirical question, which will be dealt with in subsequent chapters. A third type of estimator, the finite mixture model, is in its infancy in applications to risk-adjustment, so we postponed an examination of it until Chapter 4. We proposed a method for accounting for hierarchical nature of the dataset and discussed potential endogeneity.

Having reviewed the elements of estimating a theoretically and econometrically consistent and precise utilisation function, section 2.3 examined how the results of the utilisation function can be used to generate budgets. Two considerations were discussed in detail, namely, distributional effects and risk exposure. The review of distributional effects examined equity and unmet need. The review of risk exposure examined risk selection and risk management strategies. The concepts outlined will be measured in the empirical chapters below to inform our selection of the preferred model of budget setting.

3. DATA DESCRIPTION, QUALITY ASSESSMENT AND IMPUTATION

This chapter describes the data used to populate the model in equation (2.2), as well as additional variables used to evaluate competing models drawing on the discussion in section 2.3. It is a wide ranging chapter, offering a number of significant contributions. First, we identify a dataset on socio-economic and demographic circumstances of GMS recipients, as well as their access to GP services. This is the first time this dataset has been used for research, as far as we are aware. Second, the we generate chronic illness indicators based the chronic disease score methodology for the first time for Ireland. Third, we propose and apply a test for measurement error in the chronic illness indicators. Fourth, we produce a number of new indicators of GP prescribing style for Ireland, based on McGavock (1988). Fifth, we apply an innovative approach to dealing with missing values, involving multiple imputation. This is one of the few applications of multiple imputation in applied economics. Sixth, we impute income for every individual in the sample, which is again the first time it has been done for Ireland, and one of the very few times income has been used in the assessment of risk-adjustment models.

Demographic and socio-economic variables and indicators of access to GP services are generated from the Medical Card Register. The Medical Card Register is the database used to manage an individual's eligibility for GMS services. It is maintained in each health board separately. Applicants for the medical card are assessed for eligibility by Health Board-employed Community Welfare Officers, under a number of criteria, broadly age and income, as described in Chapter 1. Certain particulars of the eligibility assessment test are recorded onto the Medical Card Register. While the Registers differ by health board with respect to types of data collected, a core set of common variables relevant to this study have been identified. As well as date of issue of the medical card, there are variables related to GP payment – age, gender and distance to the GP – and additional universally recorded variables – disability, lone parenthood, urban/rural residence and marital status. In addition a number of variables relating to vulnerable groups are recorded on some but not all Medical Card Registers. Since they are not universally recorded we cannot include them in the estimation process. However, since the ultimate test

of each model is its ability to make unbiased, precise and fair predictions, external variables such as these are extremely useful for model evaluation. This study is the first time the Medical Card Register has been used for research, as far as we are aware, and constitutes a significant discovery in our view.

When a GMS patient gets a prescription from the GP, they cash it in either in a pharmacy or, if their GP has a dispensing license, at the GP surgery. The medicine is dispensed free of charge to the patient and a duplicate of the prescription is sent by the dispenser to the GMS (Payments) Board for payment. A number of details are entered into the GMS prescribing database, including a patient identifier, patient age, GP identifiers, dispenser identifier, expenditure data and drug details, which follow the anatomical therapeutic chemical (ATC) classification scheme. Since this dataset relates to payment of dispensers, it is of good quality. It is currently used to generate the current indicative drug targets, based on the NARA, while this study uses the dataset for three purposes.

First, we use the prescribing expenditure data as the response variable. Second, we generate chronic illness indicators, based on the chronic disease score methodology, from the drug details that the dataset captures. The review of determinants of health care utilisation in Chapter 2 highlighted the importance of health status. Risk-adjustment studies have favoured either diagnosis-based measures or measures based on prescribing data, using the chronic disease score methodology. We generate the latter. As well as generating them, we compare them to epidemiological estimates of prevalence of chronic illness, to assess their external validity. The generation of chronic disease scores is a second innovation described in this chapter. Furthermore, one of the outstanding concerns with using previous utilisation to measure health status, whether it be diagnosis data or prescribing data, is that it may contain measurement error, as described in section 2.2.2. We propose a test for measurement error and apply it to our chronic illness indicators, which is a third contribution in this chapter. Third, we use the prescribing data to generate indicators of GP prescribing style based on McGavock (1988). This is a fourth contribution in this chapter.

The GMS (Payments) Board manage each GP's GMS contract, so they also provide this study with data on other GP characteristics such as age, presence of a nurse or secretary and size of GMS panel.

We find that the number of missing values for some variables, especially disability, is quite high. While the default approach to dealing with missing values in most econometric analyses is to exclude observations with missing values, we use multiple imputation. This has been used rarely in econometrics and is a fifth innovation in this chapter.

Finally, we predict income for each individual in the study. We identify variables common to both the Household Budget Survey (2000) and the Medical Card Register. These variables are used to model income for the Household Budget Survey sample. This model of income is then used to predict income for all individuals in the Medical Card Register. Predicted income is then used in Chapters 6 and 7 as part of the model evaluation exercise, and is likely to be of use in future research on GMS recipients.

Section 3.1 outlines the inclusion criteria for the study population and discusses the harmonisation of datasets. Section 3.2 then describes the response variable and each of the covariates, including the generation of chronic illness indicators and indicators of prescribing style. Since the data are not routinely used for research, their quality is also assessed, by identifying clear errors in each variable. A new test for measurement error in chronic illness indicators is also applied. Section 3.3 compares variables with similar variables from other datasets as a check of external validity. Methods for dealing with missing values are discussed in section 3.4. Section 3.5 examines collinearity in the variables used in the utilisation model. Section 3.6 describes the variables used for model evaluation and describes the prediction of income based on the Household Budget Survey 1999/2000. Section 3.7 summarises and concludes.

Before concluding the introduction, however, a short profile of the sample is in order. We use data from three of Ireland's 11 health boards, namely, the Southern, South Eastern and North Eastern Health Boards. The sample size is 400,751, of

which 42% are from the Southern Health Board, 31% are from the South Eastern Health Board and 26% are from the North Eastern Health Board. This constituted 37% of the population of GMS recipients in 2000 and was representative of the GMS population with respect to age, gender and prescribing expenditure, which were the three variables for which we had data for a the full population. In addition, these three health boards have a mix of urban and rural areas. The country's second and fourth largest cities are in the Southern and South Eastern Health Boards respectively, while there are a number of large urban areas in the North Eastern Health Board. Meanwhile, the peninsulas in the Southern Health Board are amongst the most sparsely populated areas in the country. Regional GDP is almost exactly equal to the national average in the Southern Health Board, and is slightly lower than average in the North Eastern Health Board and South Eastern Health Board, but not remarkably so. Therefore, we are confident that the results from this sample are applicable to the entire GMS population.

3.1 INCLUSION CRITERIA

Data from the Medical Card Register are for a point in time, the date of which varies by Health Board, while expenditure data relate to a 12 month period and indicators of chronic illness are generated over a 9 month period. Chronic illness indicators are measured over the period September 1999 to May 2000 inclusive. Annual prescribing expenditure refers to the 12 months from June 2000 to May 2001 inclusive. The lag is introduced because we do not want the response variable and a covariate to be generated from the same dataset. This approach was also adopted in Hornbrook et al. (2001). Meanwhile, Medical Card Register data from the Southern Health Board refer to the 4th April, 2000; for the South Eastern Health Board they refer to 24th October 2000 and for the North Eastern Health Board they refer to 24th November 2000.

Over the study period, the GMS population changes as a result of changes in eligibility status, migration, births and deaths. Consequently, the population on the Medical Card Register does not correspond perfectly either to the population indicated as having chronic illness or the population recording prescribing expenditure during the cost year.

The implications of this lack of correspondence are, first, that those people who are GMS recipients for only part of the cost year June 2000 to May 2001 will have their annual expenditures underesetimated, while those who are GMS recipients for only part of the chronic illness estimation period September 1999 to May 2000 may not have a chronic illness from which they suffer recorded. A number of adjustments are possible to remedy these problems, at least partly. These adjustments are reliant on the validity of the date of issue variable. Two tests of its validity are described in Appendix 3.1, which concludes that it is a valid variable.

The adjustments required to harmonise datasets are described in Table 3.1.

Table 3.1[6]

Adjustments to Study Population to Harmonise Data

Health Board	No. on MCR[1]	Issue Date Coded	Issued 9/99 – 5/00	Issued 6/00 – 11/00	Born 6/00 – 11/00
NEHB	106,598	105,637	8,594	4,938	262
SEHB	135,187	125,217	11,861	4,369[3]	303
SHB	169,897	169,897	18,207[2]	-	-
Total	411,682	400,751	38,662	9,307	565

NOTE 1: MCR = Medical Card Register
NOTE 2: Up to 4th April 2000
NOTE 3: Up to 24th October 2000

The number of people on the Medical Card Register in the three health boards areas was 411,682. We find that the removal of observations who do not have the date of issue coded reduces the study population to 400,751, of which 42% are from the Southern Health Board, 31% are from the South Eastern Health Board and 26% are from the North Eastern Health Board.

This reduction in sample size from 411,682 to 400,751 is due almost entirely to missing issue dates in the South Eastern Health Board. We find in Appendix 3.1 that those missing issue date are disproportionately concentrated in the under 12s and 24-45 year age groups, while very few are over 65. This suggests that their cards were recently issued. Moreover those with the issue date missing have an average expenditure of IR£96 as against an average of IR£208 for all others, while a test for sample selection bias finds that those with missing issue dates are systematically

different from the rest of the population. However, it is reasonable to assume that the difference in expenditures is due largely to those with missing issue dates having their cards issued during the cost year and therefore having less than 12 months expenditures. The relatively small number of cards that were issued during the cost year in the South Eastern Health Board and the relative age of those with the issue date missing supports this argument. Consequently we assume that the reason that those records with missing issue dates have lower average expenditures than other records is because many of them were issued during the cost year and if a full year's expenditures were available, they may not be systematically different from the rest of the sample. As such excluding them does not introduce a sample selection bias. This assumption is, however, untestable. Appendix 3.1 includes a description of the difference between expenditures of those with issue date coded and those it missing, by age-group, as well as a description of the test for sample selection bias.

The fourth column of Table 3.1 describes the number of cards issued from September 1999 to May 2000 inclusive. This is the period when chronic illnesses were measured. If cards that were issued during the period September 1999 to May 2000 are included in the measurement of chronic illness, underestimates of prevalence rates will emerge. Therefore the 38,662 individuals who had cards issued between these dates are excluded from the measurement of chronic illnesses, meaning that they are excluded from some of the models estimated below.

Column five of Table 3.1 describes the number of cards issued during the cost year. If cards that were issued between June 2000 and May 2001 are included in the calculation of annual prescribing expenditures, those expenditures will be underestimates. This does not apply to the Southern Health Board, since the cross-section of the Medical Card Register for that board refers to April 2000, before the start of the cost year. Meanwhile there were 4,672 cards issued up to October 2000 in South Eastern Health Board (of which 303 were to new-born babies) and 5,200 up to November 2000 in the North Eastern Health Board (of which 265 were new-born babies). The expenditures recorded for these people are annualised by using the inverse of the fraction of the year that these people did not have medical cards. For

[6] Unless otherwise indicated, the data sources for all tables are the GMS(Payments) Board 1999 – 2001; the Southern Health Board, 2000; the North Eastern Health Board, 2000 and the South Eastern

instance, someone who had a medical card issued in October 2000 was present for 4/12 of the cost year. Therefore their costs are inflated by 12/8.

Obviously, any cards that were issued after the date of the Medical Card Register in each health board are automatically excluded from the analysis, as we do not have any demographic and socio-economic variables for these people. Finally, there will be people whose cards expired during the cost year. Their expenditures will be underestimated by that fraction of the cost year that they were not in the scheme. Unfortunately we cannot identify these observations.

3.2 *VARIABLE DESCRIPTION AND QUALITY ASSESSMENT*

This sub-section defines each variable used in the estimation of the determinants of prescribing expenditure and tests the reliability and validity of each one, where relevant.

As with any empirical investigation, we need to assess the reliability and validity with which the data have been measured. Measurement focuses on the link between empirically grounded indicators, the observables, and the underlying unobservable concept. When the link is a strong one, then the observables can be used to form useful inferences about relationships in the underlying concepts, while wrong inferences may be drawn in the case of a weak link between the two. In order to determine the degree to which an observation measures an underlying concept, two criteria are used – reliability and validity. Reliability refers to the extent to which an experiment, test, or any measuring procedure yields the same results on repeated trials. In the instance of the GMS database, reliability would refer to the extent to which an assessment of an individual's entitlement to a medical card would produce the same data on repeated occasions. Validity refers to the extent to which an indicator measures what it is intended to measure.

Reliability and validity are defined above in terms of the degree to which they are present, rather than their absolute presence or absence. This is because it is accepted that repeated measurement will almost never produce exactly the same set of results,

---

Health Board, 2000.

while any indicator will not measure the underlying concept perfectly well. Data will never be perfectly reliable and valid.

In assessing errors in the dataset, it is necessary to distinguish between random and non-random errors. Random error occurs due to chance in the data collection process, whereas non-random error are due to some systematic bias in the data collection process. Non-random error is crucial to the assessment of validity, since it prevents measuring instruments from measuring the underlying concept. Therefore, reliability is inversely related to the amount of random error in the data, while validity is inversely related to the amount of non-random error in the data.

The list of variables used for model estimation is outlined in Table 3.2, followed by a discussion of their reliability and validity.

Table 3.2
Variable Description

Variable	Description
Response	
Totcost	Annual prescribing expenditure (IR£)
BN	Prescribing expenditure with budget neutral drugs removed (IR£)
BNCooksD	As BN but truncated at IR£1719.38 (IR£)
Demog. / Socio-Economic / Access	
Age	Age
Agesq	Age squared
Gender	Female (= 1)
Disabil	Member of household in receipt of disability payments (=1)
Lonepare	Member of household in receipt of one parent family allowance (=1)
Marry	Marital status of head of household (married = 1)
Rural	Member of household in rural area (=1)
Dist1	Member of household less than 3 miles from GP's principal surgery (= 1)
Dist2	Member of household is 3-5 miles from GP's principal surgery (= 1)
Dist3	Member of household is 5-7 miles form GP's principal surgery (= 1)
Dist4	Member of household is 7-10 miles from GP's principal surgery (= 1)
Dist5	Member of household is 10+ miles from GP's principal surgery (= 1)
Numcard	Number of people on each medical card
NEHB	North Eastern Health Board (= 1)
SEHB	South Eastern Health Board (= 1)
SHB	Southern Health Board (= 1)
Chronic Disease Scores	
CVD	Cardio-vascular disease (=1)
Epi	Epilepsy (=1)
Rheum	Rheumatological illness, pain and inflammation (=1)
Diabetes	Diabetes (=1)
Glau	Glaucoma (=1)
Respir	Respiratory illness, asthma (=1)
Thyroid	Thyroid disorders (=1)
Psych	Psychiatric illness (=1)
Comor	Number of comorbidities
GP Characteristics	
GPage	GP age
GPagesq	GP age squared
Nurse	Practice nurse (= 1)
Secretary	Practice secretary (=1)
RPA	Rural Practice Allowance (=1)
Decpanel	Number on GP's GMS panel in December 2000
Specific	Proportion of total prescribing that is specific
Symptomatic	Proportion of total prescribing that is symptomatic
Presum	Proportion of total prescribing that is often presumptive

As well as the variables above, we also generated an indicator of people who were aged over 18 and who did not have their own medical cards, speculating that these were high cost adult dependants. A very small number of individuals were

identified. We found that they were insignificant in all models so we do not report results with this variable included.

3.2.1 Response Variable - Prescribing Expenditures

Monthly prescribing expenditure data have been provided by the GMS (Payments) Board. We aggregate to annual data. The cost year is June 2000 to May 2001 inclusive. Expenditure is calculated as 'net ingredient cost', that is, cost exclusive of VAT and dispensing fees.

As mentioned in section 2.3, the GMS includes a number of drugs that are 'budget neutral', that is, the GMS scheme covers them, but they are not included in the calculation of indicative drug targets. These are mostly high cost drugs including a number of cancer drugs, drugs used to assist in kidney transplants and lipid lowering drugs. A full list of budget neutral drugs is in Appendix 3.5. In addition, the indicative drugs target scheme moderates the risk exposure of any one GP to high cost patients by setting a ceiling on the costs attributable to the budget for any one patient. This ceiling was IR£2075 in 2000. Alternative levels of truncation are discussed in the exploratory data analysis conducted in Chapter 4, which concludes that truncation should be applied at IR£1719.38. This is the highest value of any observation that is not identified as an outlier according to Cooks D, hence the third response variable described in Table 3.2.

The preferred response variable from these three is discussed in detail in Chapter 4.

3.2.2 Covariates – demographic, socio-economic and access

As discussed in the introduction to this chapter, these variables have been generated from the Medical Card Registers in each Health Board. Since it is the first time these data have been used for research, quality assessment is an important element of data description.

We describe the construction of those covariates that are not self-evident, namely, lone parenthood, disability, marital status, distance to GP and rural residence. The lone parenthood and disability variables are constructed from data on social welfare payments and collected for the Medical Card Register. All information on

entitlement to these payments comes from the Department of Social, Community and Family Affairs (1999) supplemented by the Department of Social Welfare (1994).

In order to detect errors in the disability and lone parenthood fields, we checked that the data met the qualifying criteria for each type of social welfare payment. We also examined the data for unexpected patterns which may still be in line with the qualifying criteria, but which we would not expect. For instance, whilst it is possible for middle aged men to be collecting one parent family allowance, it is unlikely that they would form a significant proportion of total claimants. Since these variables refer to the cardholder rather than the individuals in households, we can only check that the cardholders are meeting the qualifying criteria.

There are a number of ways of adjusting the database for the identification of anomalies. When an error could have occurred in a number of different fields, it is assumed that the discretionary field has been coded incorrectly. A discretionary field, such as occupational status, is one where the data are not collected for payment purposes, and is therefore more susceptible to poor coding than compulsory fields such as age. If we find that an anomaly exists and this could be due to a coding error in a discretionary or a compulsory field, then we always assume that it is the discretionary field. We then insert a missing value for this field.

3.2.2.1 Lone Parenthood

The lone parenthood variable is generated from data on social welfare payments to one-parent families. While there is now only one payment made to one-parent families, there used to be a number and our database includes people coded under the old payment schemes. First, what is now called One Parent Family Payment covers what used to be known as Lone Parent Allowance, Maintenance Allowance, Deserted Wife's Allowance and Deserted Wife's Benefit. These older names appear in the Medical Card Register in some cases.

One Parent Family Payment is payable to you if you are bringing up a child or children without the support of a partner. You qualify if you are widowed, separated, deserted, unmarried or a prisoner's spouse and if you support at least one child and are not cohabiting. It is means tested. Gross earnings must be IR£12,000 per year or

less for full entitlement. The first IR£6,000 per year of your earnings is disregarded and half of the remainder of your earnings up to IR£12,000 per year is tested as means. If your subsequent earnings exceed IR£12,000 per year, you are paid half of your Lone Parent's Allowance for a period of one year.

You are defined as deserted and qualify for this payment if the following conditions are met: you have been deserted by your husband for at least three months; you have made appropriate efforts to get maintenance from him; your husband is not paying adequate maintenance for you and your children; you are aged 40 or over and have no dependent children (this age limit does not apply if you have dependent children); you are not cohabiting and you satisfy a means test.

As a check for anomalies, we expect that the majority of cardholders coded as recipients of lone parent payments are women of childbearing age. We find that 95% of the cardholders in this group are indeed women. However, there are 132 cardholders who are under thirteen, some of whom are babies. Of these, 69 are male. This suggests that the under 13s are probably errors and we code them as missing. There are also 131 cardholders who are over 66, all of whom are females. These are unlikely to be collecting this allowance, since they are unlikely to have dependants, and at any rate, they should be receiving an old age pension as well. Again, these are likely to be errors and are coded as missing. In total, 0.1% of observations are coded as missing in the lone parenthood field due to detected anomalies. We can only test for logical inconsistencies, however. A prospective examination of the data collection process would provide further information on the validity of this indicator.

3.2.2.2 Disability

This variable consists of three types of social welfare payment. First, there is Disability Allowance (formerly Disabled Person's Maintenance Allowance). This is a weekly payment administered by the local Health Board paid to people with a disability who are aged over 16 and under 66 and not in full-time residential care. The allowance is granted subject to medical suitability and a means test. Medical suitability is defined as having an injury, disease, illness or physical or mental disability which has continued or can be expected to continue for at least one year. As a result of the condition you are 'substantially handicapped' in undertaking work

of which you would otherwise be capable. Medical suitability is determined by a Medical Assessor of the Department of Social, Community and Family Affairs. Disability Allowance lasts for as long as you satisfy the qualifying conditions.

Second, Disability Benefit is paid to insured people who are unfit to work due to illness. Qualifying conditions are that recipients are under 66 years of age, unfit to work due to illness and satisfy the national insurance contributions conditions. In order to claim this benefit, a medical certificate must be sent weekly to the recipient's local Social Welfare Office.

Third, Invalidity Benefit is paid instead of Disability Benefit if you have been unable to work for more than 12 months. In certain cases of serious incapacity, this benefit is also available to people who have been receiving Disability Benefit for less than 12 months. In order to qualify, you must be permanently incapable of work and satisfy the national insurance contribution conditions. The payments last as long as you meet the qualifying conditions. Invalidity Benefit can be treated as the national insurance equivalent of Disability Allowance.

In order to test for anomalies in the dataset, we check for observations that do not meet the qualifying conditions. Claimants of disability allowance must be aged between 16 and 65. We find that all cardholders are in this age group. Whilst disability benefit is payable to under 66s only, invalidity pension is payable to all ages. Therefore the only test we can apply is that they are old enough to have made national insurance contributions, that is they are more than 16 years and 9 months. We find that only 16 are less than 16 years and 9 months. These 16 are coded as missing.

What is of interest about these categories, in addition, is that they are male dominated. While we expect more of those cardholders entitled to national insurance benefits to be male, we find that in these categories it is far more dominant. Seventy two per cent of the cardholders in receipt of disability benefit and invalidity pension are male, as opposed to 55% of those entitled to disability allowance, and only 51% of all cardholders.

In conclusion, and as with the lone parenthood variable, a tiny number of anomalies were detected. The male dominance of the disability benefit / invalidity pension group is an interesting finding, but we do not consider that it questions the validity of the variable. As with the lone parent indicator, we can only test for logical inconsistencies, and a prospective examination of the data collection process would be welcome.

3.2.2.3 Marital Status

There are six categories of marital status – married, single, widowed, divorced, separated and cohabiting. However, presumably owing to a social stigma associated with divorced, separated, widowed and cohabiting people, people in these categories often get classified as single[7]. For instance, all cardholders claiming the widows pension should be widowed. We find, however, that only 32% are coded as widowed, with 66% coded as single. Therefore we reduce these six categories down to two – married and unmarried.

3.2.2.4 Distance

Payment to GPs varies with distance from the patient to the GP, since the greater the distance the greater the cost to the GP of home visits. Therefore, distance is recorded on the GMS database as distance from the patient's residence to the GP's principal surgery. While this may reflect the distance a GP travels for a home visit, it may overestimate the true distance for the patient to travel for a consultation, since many GPs operate one or more branch surgeries. Nevertheless, it offers useful information on access to GP services for the GMS population.

3.2.2.5 Rural Residence

Medical cardholders are classified as rural or urban based on local authority administrative area. No definition of urbanisation exists, so Community Welfare Officers use their judgement, which may slightly bias this variable.

3.2.3 Covariates – Chronic Disease Scores

Chapter 2 highlighted the importance of health status as a determinant of health care utilisation. By far the most popular measure of health status included in risk-

adjustment studies are either diagnosis-based measures or measures based on prescribing data, using the chronic disease score methodology. We use the GMS (Payments) Board prescribing dataset to generate the latter.

Discussion of chronic disease scores is divided into two parts. First, we describe the chronic disease score methodology and modifications to it for this study. Then we undertake a detailed assessment of its validity as a measure of morbidity, including the application of a new test for measurement error in chronic illness indicators. Section 3.3 includes a comparison with epidemiological estimates of prevalence of chronic illness, to assess their external validity.

3.2.3.1 Modifying Chronic Disease Score Methodology for Irish Setting

The chronic disease scores methodology constructs indicators of chronic illness based on prescribing data. The original CDS (Von Korff et al., 1992), created by identifying chronic diseases using pharmacy claims, was found to have construct and predictive validity (Johnson et al. 1994). A revised CDS (Clarke et al., 1995), covering a wider range of medicines, produced a set of 28 dummy variables identifying a prescription for a particular medication or medication class which indicates the presence of a chronic condition. A modified version of this CDS was used by Lamers (1999a) and Lamers and van Vliet (2001) for risk-adjustment in the Netherlands and Hornbrook et al. (2001) in the US.

A number of GPs in Ireland and the UK were interviewed for their opinion on each drug included in each CDS category. Consequently, there are a number of modifications to Lamers' (1999a) scheme. The first modification arises out of a concern that some drugs included in the scheme reflected the prescriber's attitude to prescribing as much as underlying morbidity. In order to overcome this problem, we decide that if McGavock (1988) classified a drug as other than 'specific' (that is 'symptomatic' or 'often presumptive'), then we remove it from our list of chronic disease indicators. If McGavock classified a drug group as other than 'specific', then we examine drugs within that group to ensure that they are no 'specific' drugs within a drug group that is primarily not 'specific'. Any 'specific' drugs are retained and the rest are excluded.

---

[7] Personal communication, Southern Health Board

In addition, in order to keep our model as parsimonious as possible, a number of similar chronic diseases are grouped together. Furthermore, any indicators that are rarely used in General Practice are excluded. All of the excluded indicators had a prevalence of much less than 1% of the GMS population. Appendix 3.2 illustrates the link between the chronic disease scores in Lamers (1999a) and those used in this study and identifies those drugs excluded because McGavock (1988) did not classify them as 'specific'.

3.2.3.2 Chronic Disease Scores as a measure of Morbidity

Section 2.2.2.2.2 discussed the validity of previous utilisation of health services as a measure of morbidity. We found that three concerns arose about the chronic disease score methodology. First, multiple indications for some medicines existed. We found that the problem of multiple indications is circumvented by using conservative inclusion criteria. Second, incidental users needed to be distinguished from chronically ill. Third, chronic disease indicators only measured morbidity of a number of conditions held, relating principally to access to services and physician prescribing style. We deal with the second and third points in turn.

In order to distinguish between incidental users and the chronically ill, different studies apply different inclusion criteria. Lamers (1999a) identifies someone as having a chronic illness if they have four or more prescriptions of the relevant medicine a year. Hornbrook et al. (2001) indicate anyone who has one or more, so their inclusion criteria are much more lax. Lamers and van Vliet (2001) identifies a chronic illness if more than 181 defined daily doses (DDDs) have been prescribed in one year (Lamers and van Vliet, 2001). While a defined daily dose field is indicated on our prescribing dataset, it appears to be unreliable for certain medicines. Therefore, we identify someone with a chronic illness if they have at least one prescription of a relevant medicine every three months over the period September 1999 to May 2000. As such, it is stricter than either Lamers (1999a) or Hornbrook et al. (2001), although probably not as conservative as Lamers and van Vliet (2001).

The literature provides no clear solutions to the potential for chronic disease scores to be biased measures of morbidity. We propose to specify it as a measurement error problem. The problem is described using the following utilisation function:

$$y_{ij} = \beta x_{ij} + \gamma h_{ij} + \theta z_j \qquad (3.1)$$

where y_{ij} is prescribing expenditure of individual i attending GP j; x_{ij} is a vector of well measured individual covariates; h_{ij} is health status measured using chronic disease scores and z_j is a vector of well measured supply characteristics. Suppose health status is measured with error as follows:

$$c_{ij} = h_{ij} + u_{ij} \qquad (3.2)$$

where c_{ij} is our measure of health and u_{ij} is the error term and u_{ij} is unrelated to h_{ij}. Therefore,

$$y_{ij} = \beta x_{ij} + \gamma c_{ij} + \theta z_j - \gamma u_{ij} + e_{ij} = \beta x_{ij} + \gamma c_{ij} + \theta z_j + \tau_{ij}, \qquad (3.3)$$

where τ_{ij} is a residual that mixes the measurement error and the 'true' error. The estimates of γ, which we refer to as g, will be downward biased from the true estimates, because c_{ij} is negatively correlated with τ_{ij}. This result only holds if u_{ij} is uncorrelated with e_{ij}, which is reasonable. Bounds can be placed about γ by running a reverse regression, c_{ij} on y_{ij}, x_{ij} and z_j. The inverse of the coefficient on y_{ij}, l, is the upper bound of γ (Hausman, 2001: 59), so $g<\gamma<l$. However, g/l equals the R^2 (of either the regression or reverse regression, as they have the same R^2), so given an average R^2 of about 0.3 for the above utilisation equation, l is three times g and the bounds are very wide.

If one is suffering from a chronic illness, individual factors such as the disutility associated with the process of health care, understanding of illness and access to GP services are unlikely to present meaningful barriers to that disease being presented to the GP. We present results in Appendix 3.3 to support this contention, as part of the analysis that follows. Neither is it likely that patients fail to cash in their prescription routinely. Thus, the principal source of measurement error of h_{ij} is the practice style of the GP. The prevalence of a particular chronic illness in a GP's practice list will be due in part to the GP's diagnostic ability, which we assume varies across GPs for that condition, and across conditions for any one GP. Hence:

$$u_{ij} = f(z_j). \qquad (3.4)$$

There are a number of ways of dealing with measurement error, the most common being the use of instrumental variables. In this context, there are no variables that are correlated with c_{ij} but unrelated to y_{ij} – any variable available in the study that predicts chronic illness also predicts prescribing expenditure and should be included in the utilisation function. Two stage least squares is still an option however.

We apply the following as a first stage regression for each chronic illness indicator (with subscripts suppressed):

$$\text{Prob}[C = 1] = \Phi(\lambda x), \qquad (3.5)$$

where C is one if the individual has the chronic illness and zero otherwise, x is a row vector of covariates, λ is the set of parameters to be estimated and $\Phi(.)$ is the standard normal distribution.

In the problem to hand we have 'weak instruments', meaning that these first stage regressions have poor explanatory power and the instrumental variables test does not detect mismeasurement as the standard errors are imprecise. These are illustrated by the low level of explanatory power of some of the first stage regressions in Table 3.3.

Table 3.3

Explanatory Power of Prediction Models of Chronic Illnesses

Chronic Illness	Pseudo-R^2 (%)
Cardiovascular disease	29.49
Epilepsy	8.31
Rheumatology	9.60
Diabetes	10.54
Respiratory Illness	3.28
Glaucoma	16.92
Thyroid Illness	13.47
Psychiatric Illness	12.09

Although explanatory power is reasonable for cardiovascular disease and, arguably, for glaucoma, thyroid illness and psychiatric illness, it is poor for respiratory illness, epilepsy and rheumatology. Unless the instrumental variables are jointly strong enough, alternative estimation strategies should be adopted. As Verbeek (2000: 129) states: *"Mainly due to the problem of finding suitable instruments, the problem of measurement error is often ignored in empirical work"*.

Therefore we propose an alternative. We test the effect of differences in practice style on measures of health status is measured as follows. First we regress each chronic illness indicator on exogenous individual characteristics and GP practice style variables, as per (3.5). Predicted probabilities of chronic illness are generated from this model. Then predicted probabilities are generated where the GP practice style variables are set at the sample average. Differences in predicted probabilities are then assessed, giving an indication of the extent of bias.

Differences in predicted probabilities only have a substantive effect if they change the indicator variable from 0 to 1 or vice versa. We assign all observations with a predicted probability of greater than or equal to 0.5 a value of 1 and those with a predicted probability of less than 0.5 a value of 0. Differences in chronic illness estimates are then compared[8].

To reduce computational time, we use a random sample of 50% of the dataset to generate estimates of the probability of having each chronic illness and predict probabilities on this basis. Table 3.4 illustrates the deviation in predicted probabilities when supply variables are allowed to vary for prediction and when they are frozen. The prediction equations from (3.5) for each chronic illness are outlined in Appendix 3.3.

Table 3.4

Prediction of Chronic Illness: Deviation in Predicted Probabilities (%) between Prediction Strategies

Percentile	CVD	Epilepsy	Rheumatology	Diabetes	Respiratory	Glaucoma	Thyroid	Psychiatric
5%	-4.6	-0.4	-0.6	-0.6	-1.4	-0.2	-0.6	-1.5
25%	-2.0	-0.1	-0.1	-0.1	-0.3	0.0	-0.1	-0.2
50%	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0
75%	1.0	0.1	0.2	0.1	0.5	0.0	0.1	0.4
95%	5.5	0.5	0.7	0.8	1.2	0.4	0.7	1.3
Mean	0.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0

With the exception of cardiovascular disease, the difference between prediction based on the full prediction model and one where the supply variables are frozen is

[8] An alternative to setting values of greater than 0.5 would be to assume that the predicted probability is from the binomial distribution and translating a draw from the underlying binomial distribution back to either 0 or 1 (Honaker et al., 2001).

negligible. The mean difference is zero for all except cardiovascular disease and respiratory disease, while the difference between the 5th percentile and 95th percentile is less than 1% for epilepsy and glaucoma, less than 2% for rheumatology, diabetes and thyroid problems, less than 3% for psychiatric and respiratory illness. The deviation for cardiovascular disease just over 10% however, which is a little more noteworthy.

Given these differences in predicted probabilities, the number of individuals that are mis-classified in chronic illness due to the prescribing style of their GP is likely to be very small indeed. The one exception is cardiovascular disease where a number of individuals may well be misclassified. Unlike the other chronic disease areas, a number of cardio-vascular drugs are likely to be associated with prescribing style. For instance, high cholesterol is a risk factor for heart disease and cholesterol modification therapy is recommended if a patient is considered at 'high risk' of heart disease, that is if they have a number of other risk factors as well as high cholesterol. However, the GP has some discretion is deciding if a patient is 'high risk'. In addition there are alternatives to pharmacotherapy to reduce cholesterol. Therefore, it is not particularly surprising that of the chronic illness indicators presented above, cardio-vascular disease has the strongest relationship with prescribing style.

Overall, this analysis, which supports the findings of Ellis (2002), finds that measurement error due to the effects of practice style on measurement of chronic illness is likely to be quite small. Correcting for measurement error, especially in the presence of weak instruments, is unwarranted. Although the measurement error in cardiovascular disease is still not huge, its presence should be borne in mind when interpreting the results. Future research should consider alternative strategies to correct for measurement error or develop better instrumental variables. Finally modification of the drugs included as indicators of cardio-vascular disease in the chronic disease score methodology should be considered.

The chronic disease scores produce data on health status for everyone in our study population, for a recent period of time. They have been used in weighted capitation models in the Netherlands (Lamers and van Vliet, 2001) and the US (Hornbrook et al., 2001) where they have outperformed alternatives. Although they have some

drawbacks, we have taken reasonable steps to test their validity. In particular, we applied a test of measurement error that suggested that GP prescribing style had a very small effect on the indication of someone as a chronic illness sufferer. This may be of interest to other researchers in the field, and is a useful empirical complement to Ellis' (2002) theoretical work.

3.2.3.2 Comorbidities

Previous studies using chronic disease scores for risk-adjustment have not separated individuals who have only one chronic illness from those who have more than one, that is, who have comorbidities (Fishman and Shay, 1999), with one exception (Lamers and van Vliet, 2001). However Lamers and van Vliet (2001) took the extreme approach of allocating someone to only one chronic illness no matter how many they had. Thus they did not allow for comorbidity at all, which is likely to increase prediction error for their high cost patients. Since chronic disease scores do not indicate the primary illness, Lamers and van Vliet (2001) allocated the individual their most expensive chronic illness only.

We consider two specifications of comorbidity. First we apply interactions between each chronic disease indicator. Second, we include a comorbidities variable that indicates the number of comorbidities an individual had as outlined in Table 3.5.

Table 3.5

Prescribing Expenditure by Number of Chronic Illnesses

Number of Chronic Illnesses	Average Prescribing Expenditure	Marginal Increase in Prescribing Expenditure
0	82.13	
1	469.16	387.03
2	770.71	301.55
3	1065.29	294.58
4	1332.32	267.03
5	1796.77	464.45
6	184.31	-212.46

Each additional chronic illness up to the fifth is associated with an average increase of between IR£267.03 and IR£464.45. These are similar in magnitude, if slightly lower, to the coefficients on chronic illnesses. Therefore it appears that chronic illnesses can be added linearly, while the coefficient on the comorbidities variable using the second approach is likely to be negative but small, reflecting the

diminishing marginal increases in expenditure. In addition, some of the interactions have very low prevalence leading to unstable coefficients. Therefore, we use the second approach to model comorbidity.

3.2.4 Covariates - GP Characteristics

GP characteristics data are available from the GMS (Payments) Board, for both indicators of prescribing and other GP-related indicators. The prescribing style variables relate to the period August 1999 to May 2001 (the same period as the chronic disease score measurement) while other GP-related indicators were provided by the GMS (Payments) Board in October 2001. GP age, nurse, secretary and panel size variables are self-evident. Rural Practice Allowance is paid to GPs who practice is particularly remote areas.

Prescribing style is measured using McGavock's (1988) teleological classification of drug types reflecting perceived use in general practice. As described in section 2.2.4.2, drugs types are categorised as 'specific', 'symptomatic' and 'often presumptive'. The average percentage of total prescribing that is specific is 48%, the average percentage that is symptomatic is 20% and the average percentage that is often presumptive is 17%. The classification by therapeutic group and ATC code is described in Appendix 3.2. The generation of prescribing style variables is a novel contribution of this study.

A number of tests of the reliability of the GP data are run. First, we have two figures for a GP's panel size, one for December 2000 and one for October 2001. If there is a large difference between the panel size for these two months, this suggests that the GP in question has moved, retired or is in some way 'unusual'. Specifically, any GP whose October 2001 panel is more than 10% greater or less than their December panel (as a percentage of the December panel) is dropped. Twenty six percent of GPs were discovered to have a 10% or greater change in panel size in this 10 month period, although this related to only 14% of patients, as those GPs who had large changes in panel size had below average panel sizes in most cases. For other GPs, panel sizes did not vary much, with a median change of 1.2%. Panel changes did not differ by health board area.

A second problem with the GP-level data is the reliability of the GP indicator in the prescribing database. This problem relates only to the three prescribing style variables, since these are the only ones that use the prescribing database to generate GP-level variables. All other GP level variables are generated from the Medical Card Register. As described above, the GMS (Payments) Board records details of the GP on its prescribing database for every claim that it processes. In particular, the GMS number of the GP with whom they are registered (known as the GP of choice) and the prescribing GP is recorded. For now, let us call the GMS prescribing database the patient file. We can aggregate each individual's expenditure by their GP to get their GP's total expenditures. We also received a second source of GP-level expenditures from the GMS (Payments) Board, relating to budgets and expenditure for the year 2000. We refer to this as the GP file. Obviously, aggregation of expenditure in the year 2000 by GP for the patient file should equal expenditure for that GP according to the GP file.

In fact, two inconsistencies were detected. First, the number of GPs included in the two databases differed. There were 191 GPs indicated as GPs of choice in the patient file that are not in the GP file, while there were 18 GPs in the GP file indicated as GPs of choice not indicated in the patient file, with 608 GPs common to both datasets. The 191 GPs who do not appear to be GPs of choice on the GP file may be locum GPs, who should have been coded as prescribing GPs, not GPs of choice, or other such coding errors. There is some evidence to suggest that they are locum GPs. Their average prescribing is only IR£4,658 as against over IR£120,000 for GPs who are common to both files. Indeed, their median prescribing expenditure is only IR£86.40. Therefore, these 191 GPs prescribing expenditures amount to only IR£889,678 or 1.1% of total expenditure. The 18 GPs who are in the GP file but not in the patient file recorded expenditures of only IR£536,796 or less than 0.7% of total expenditure. Therefore, these differences appear to be unimportant.

A second inconsistency is that the expenditures differed considerably between the patient file and the GP file. In personal communications with the GMS (Payments) Board, they suggested that in the data we received the GP of choice field and prescribing GP field were the wrong way around. An aggregation of patient expenditures by prescribing GP suggested that this was not so, except in certain

cases. However, aggregation by prescribing GP produced a much closer agreement with the GP file than did aggregation by GP of choice.

A comparison of prescribing expenditures according to the GP file, the patient file aggregated by GP of choice and the patient file aggregated by prescribing GP, by health board are described in Table 3.6.

Table 3.6

Difference between Prescribing Expenditures on GP File and Patient File, by Aggregation Method

Statistic	North Eastern	South Eastern	Southern	Total
Aggregation by GP of Choice				
Median Difference in Expenditure	2.5%	3.9%	5.4%	4.3%
GPs with < 10% Difference in Expenditure	70	89	187	346
% of GPs with < 10% Difference in Expenditure	53%	47%	66%	57%
Aggregation by Prescribing GP				
Median Difference in Expenditure	1.0%	4.5%	4.6%	3.8%
GPs with < 10% Difference in Expenditure	109	146	258	513
% of GPs with < 10% Difference in Expenditure	79%	74%	89%	85%

While the median difference in expenditures is only 4.3% on average, only 57% of GPs expenditures differ between the patient file and the GP file by less than 10%. Indeed in the South Eastern Health Board a majority of GPs expenditures differ by more than 10%, with only 47% having differences of less than 10%. In the North Eastern Health Board only 53% of GPs have differences in expenditure of less than 10% while in the Southern Health Board 66% of GPs have differences of less than 10%. The results of aggregation by GP of choice bear scant relationship with their reported expenditures according to the GP file. As such, we conclude that the GP of choice field in the patient file is unreliable.

Differences are not so great when expenditures are aggregated by prescribing GP. In this case, median differences are 3.8% and 85% of GPs prescribing expenditures are no more than 10% different between the patient and the GP files. This ranges from 74% in the South Eastern Health Board to 89% in the Southern Health Board.

If the fields were the wrong way round, as the GMS(Payments) Board believed, then differences should be zero. If this is, in fact, the prescribing GP field, then we expect

aggregation by prescribing GP to differ from the GP file, since GPs often prescribe to patients outside their panel, especially during out-of-hours. Therefore, the results of aggregation by prescribing GP are consistent with expectations. This field is used to construct the three prescribing style variables only. If it represents the prescribing GP, then these variables are measuring the prescribing style of that GP for all patients they see, both from their own panels and from other GP's panels. If the field represents the GP of choice field but is measured with error, then these variables equally are measured with error.

3.2.5 Descriptive Statistics

Having defined each variable and assessed their reliability and validity, descriptive statistics are presented in Table 3.7.

Table 3.7

Characteristics of GMS Population[1]

Variable	N	Mean	S.D.	Min.	Max.
InfNIC	400,751	211.42	450.57	0.00	17,202.33
BN	400,751	196.17	420.60	0.00	16,707.29
BNCooksD	400,751	184.30	338.48	0.00	1,719.35
Age	400,751	43.45	26.56	0.00	105.21
Agesq	400,751	2593.24	2396.15	0.00	11,069.17
Gender	400,751	0.55	0.50	0	1
Disabil	400,751	0.12	0.31	0	1
Lonepare	400,751	0.10	0.30	0	1
Marital	400,751	0.48	0.50	0	1
Rural	400,751	0.61	0.49	0	1
Numcard	400,751	2.43	1.76	1	14
0-3 miles	400,751	0.62	0.49	0	1
3-5 miles	400,751	0.14	0.35	0	1
5-7 miles	400,751	0.20	0.40	0	1
7-10 miles	400,751	0.03	0.18	0	1
10+ miles	400,751	0.01	0.12	0	1
NEHB	400,751	0.26	0.44	0	1
SEHB	400,751	0.31	0.46	0	1
SHB	400,751	0.42	0.49	0	1
CVD	362,072	0.17	0.38	0	1
Epi	362,072	0.01	0.12	0	1
Rheum	362,072	0.02	0.15	0	1
Diabetes	362,072	0.02	0.14	0	1
Glau	362,072	0.01	0.10	0	1
Respir	362,072	0.05	0.22	0	1
Thyroid	362,072	0.02	0.14	0	1
Psych	362,072	0.07	0.25	0	1
Comor	362,072	0.11	0.38	0	5
GPage	305,887	50.66	7.13	31.58	70.60
Gpagesq	305,887	2617.64	734.62	997.02	4,984.11
Nurse	305,887	0.69	0.46	0	1
Sec	305,887	0.92	0.28	0	1
RPA	305,887	0.13	0.34	0	1
Decpanel	305,887	928	341	1	1,783
Specific	305,887	0.48	0.06	0.00	1.00
Symptomatic	305,887	0.20	0.03	0.00	0.33
Presum	305,887	0.17	0.03	0.00	0.28

NOTE 1: Some values have been multiply imputed, as described in Section 3.4, amounting to 14.5% of disability values and no more than 2% of any other variable.

Prescribing expenditure varies from IR£0 to IR£17,202.33, with a mean of IR£211.42. However, once budget neutral drugs are removed the mean drops to IR£196.17, while once high cost patients are removed, average expenditures fall to IR£184.30. The reasoning behind truncation and the appropriate response variable to use are discussed in more detail in Chapter 4.

Average age is 43, varying from newborn babies to someone who is 105 years old. Risk-adjustment studies in the UK have often considered the problem of 'list inflation' where GP list (panel) sizes are exaggerated by dead people or transfers

remaining on a GP's list after the event (Scottish Executive, 1999). The UK solution is to adjust GP practice data for list inflation using the census of population. This often reduces list sizes and average ages, as those removed are usually older than average. This approach is not available in Ireland, since the census of population covers all the population whilst the GMS covers approximately one third. However, unlike in the UK, since GMS eligibility is means tested, GMS lists are updated more regularly, meaning that 'list inflation' is less of a problem. Moreover, the current IDTS formula makes no adjustment for list inflation.

We find that 56% of the population is female, 44% is male. This may seem initially like an unusually high weighting towards the females, but when one considers that more elderly receive the medical card and women live longer than men do, then more women are expected to receive the medical card. In addition, females have a higher risk of living in poverty than males, which explains their relatively greater prevalence in the sample.

Twelve percent come from households in receipt of disability payments although some of these are multiply imputed observations; 10% come from households in receipt of one parent family payments and 48% come from households where the head is married. While 61% of households are in rural areas, 61% are less than three miles from their GP's principal surgery. The percentage drawn from each health board varies from 26% in the North Eastern to 41% in the Southern.

Although there are 400,751 people in the entire sample, only 362,072 were on the GMS scheme for all of the chronic disease score measurement period (September 1999 to May 2000) and on the Medical Card Register. Seventeen percent of the sample are indicated as having cardio-vascular disease. Seven percent are indicated as suffering from psychiatric illnesses, while five percent have asthma or other respiratory illnesses. One percent of the sample have epilepsy and two percent have diabetes. The average number of comorbidities is 0.11, varying from zero to five.

Once those observations who are registered with GPs whose panel sizes changed by more than 10% from December 2000 to October 2001, and those for whom we do not have chronic disease scores, are removed, the sample size drops to 305,887. The

average patient's GP is aged about 50, varying from about 30 to about 70. Sixty nine percent of the population attend GPs who have a practice nurse, while 92% of patients attend GPs who have a secretary. Thirteen percent of patients attend GPs who get rural practice allowances, while average panel size in December 2000 was 928.

3.3 VALIDATION USING OTHER DATA SOURCES

In order to further validate the Medical Card Register, we compare the prevalence of a number of the socio-economic indicators in the Medical Card Register to similar indicators from the Census of Population in 1996 and the Household Budget Survey, 1995. In addition, the chronic disease score estimates of prevalence are compared to epidemiological estimates for the UK. As such, their usefulness not just for risk-adjustment but for other studies that require measures of community morbidity is assessed.

3.3.1 Socio-economic Indicators

The Census of Population has a comprehensive range of socio-economic indicators and can be analysed at health board level. However, it does not identify GMS recipients. The Household Budget Survey does identify GMS recipients, so a direct comparison with the Medical Card Register is possible, although not at sub-national level.

Although disability is measured in the Census as those who report having a permanent illness, we expect the GMS population to differ from the non-GMS population with respect to this indicator, so a comparison of the Medical Card Register with the Census is not enlightening. In the Household Budget Survey the number of household heads unable to work due to permanent incapacity is recorded. Although this differs from the definition used for disability in the Medical Card Register, the data should be reasonably similar. Table 3.8 compares the percentage of GMS cardholders[9] aged over 15 in receipt of disability payments with the percentage of the of households in receipt of GMS services headed by someone who

[9] Note that we refer to cardholders only, not all members of households in receipt of disability payments.

is unable to work due to permanent incapacity according to the Household Budget Survey 1995.

Table 3.8

Households headed by Disabled: GMS and Household Budget Survey compared

Health Board	Medical Card Register	HBS 1995
North Eastern	9	-
South Eastern	8	-
Southern	15	-
Total	11	8

The Household Budget Survey sampled 319 households that were headed by someone who was unable to work due to a permanent incapacity, 8% of the total GMS population sampled. The prevalence of disability on the Medical Card Register varies from 8% in the South Eastern to 15% in the Southern Health Board or 11% of average.

Table 3.9 compares prevalence of households headed by lone parents across the three surveys.

Table 3.9

Lone parents: GMS, Household Budget Survey and Census compared (%)

Health Board	Medical Card Register	HBS 1995	Census 1996
North Eastern	6	-	10
South Eastern	8	-	10
Southern	7	-	10
Total	7	9	10

There are more people coded as lone parents in the CSO publications than in the Medical Card Register, which is unexpected. The difference with the Census is especially unusual, since we expect more lone parents to be in the GMS scheme than not. However there are also more lone parents in the Census than in the Household Budget Survey's GMS sub-sample, which is also unexpected.

Table 3.10 examines the number of households headed by married people across the three surveys.

Table 3.10

Marital Status: GMS, Household Budget Survey and Census compared (%)

Health Board	Medical Card Register	HBS 1995	Census 1996
North Eastern	34	-	40
South Eastern	37	-	41
Southern	25	-	40
Total	31	29	40

Census data contains those who were ever married, except those widowed, so it includes number divorced, whereas the Medical Card Register consists of those who are married only. There are a smaller number of households headed by married people on the GMS scheme than according to the Census. This accords with the Household Budget Survey, 1995, which found that 56% of its sample were non-GMS and headed by a married couple as against 29% of its sample which were GMS recipients and from households headed by a married couple. The Medical Card Register data reports a very similar rate to the Household Budget Survey.

Table 3.11 compares the percentage of the population living in rural area according to the three surveys.

Table 3.11

Rural Residence: GMS, Household Budget Survey and Census compared (%)

Health Board	Medical Card Register	HBS 1995	Census 1996
North Eastern	57	-	58
South Eastern	65	-	59
Southern	61	-	46
Total	62	38	54

As stated above, there is no formal definition of rurality in the Medical Card Register. It is up to Community Welfare Officers to use their judgement. Meanwhile, in the Census, 'rural' is defined as settlements of less than 1500 people as well as open countryside. Household Budget Survey found that there was a higher percentage of GMS recipients living in rural areas (defined as open countryside) than non-recipients of GMS – 34% v 38%. The difference between the Medical Card Register and the Census is greatest in the Southern Health Board, while differences in the South Eastern and North Eastern are not especially great. Results suggest that Community Welfare Officers are coding areas as rural that the CSO would code as urban, especially in the Southern Health Board.

In conclusion, we found differences between the Medical Card Register and the CSO publications. These differences were never particularly large, however. This provides useful evidence supporting the Medical Card Register as a valid source of socio-economic data.

3.3.2 Chronic Disease Scores and Epidemiological Information

We wish to compare our estimates of chronic illness based on chronic disease scores with epidemiologically-derived estimates. A nationally representative health survey of the general population has never been conducted in Ireland. Estimates of diseases prevalence are mostly confined to special groups, such as those discharged from hospital, which are not representative of the general population. Thus we compare the chronic illness indicators to the Scottish Health Survey 1998 (Shaw et al., 2000) and the 1998 Health Survey for England (Erens and Primatesta 1999), which are both health surveys of the general population. If we find that they are reasonably similar, it suggests that chronic illness indicators could be used to estimate community morbidity in Ireland, in the absence of a general population health survey.

In the UK surveys 'any cardiovascular disease' included angina, heart attack, stroke, diabetes, heart murmur, abnormal heart rhythm and 'other' heart trouble. The UK health surveys use the General Health Questionnaire (GHQ12) to measure psychiatric morbidity. This is a 12 question survey covering general happiness, depression, anxiety and sleep disturbance over the past 4 weeks. A score of more than four is taken as identifying potential psychiatric illness.

Table 3.12 describes prevalence estimates of cardiovascular disease by age and gender.

Table 3.12

Prevalence of Cardio-vascular Disease in Ireland, England and Scotland (%)

Chronic Illness	16-24	25-34	35-44	45-54	55-64	65-74	16-74
MEN							
Ireland	*0.4*	*1.3*	*4.0*	*12.5*	*27.0*	*38.1*	*16.4*
Scotland	4.4	11.3	15.6	27.9	45	57.9	23.6
England	7.3	13.5	17.1	27.6	44	51.8	25.5
Northern England	7	15.4	15.8	22	49.1	49.5	24.7
WOMEN							
Ireland	*0.3*	*1.1*	*3.7*	*13.1*	*28.4*	*40.2*	*16.8*
Scotland	6.4	9.7	15.8	25.6	44.8	50.7	23.5
England	9.3	12.9	15.9	24.6	37.6	54.3	24
Northern England	8.8	12.3	18.1	21.1	36.7	58.2	23.4

Source: GMS (Payments) Board, 1999-2001; Shaw et al., 2000; Erens and Primatesta, 1999.

The prevalence of cardiovascular disease in Ireland in men aged 16 to 74 is 16.4% based on chronic disease scores. This compares with a prevalence of 'any cardiovascular disease' of 23.6% in Scotland, 25.5% in all of England and 24.7% in the North of England. The prevalence of cardio-vascular disease in Ireland in women aged 16 to 74 is a little higher at 16.8%. In Scotland, England and the North of England prevalence estimates of 'any cardiovascular disease' are 23.5%, 24% and 23.4% respectively.

As expected prevalence estimates are highly correlated with age. In Ireland, less than 5% of those aged less than 45 have coronary heart disease, which rises to 12.5% of men and 13.1% of women aged 45 to 54. Prevalence peaks amongst those aged 65 to 74 for both men and women. In Scotland, England and Northern England a similar pattern of increasing prevalence with increasing age is exhibited.

For all ages there is a greater prevalence in the UK data than in the Irish data. The Irish prevalence as a proportion of the UK measures narrows with increasing age, however. For those aged 35 to 44, for instance, the Irish measure is 25% of the Scottish measure for men and 23% for women, while for those aged 65 to 74, the Irish measures are 66% and 79% of the Scottish measures for men and women respectively. If the epidemiological estimates are more accurate, this suggests that the chronic disease scores are better at measuring coronary heart disease prevalence with increasing age. This is an unsurprising finding. For those aged less than 45 with coronary heart disease, much of it – such as high blood pressure - may go undetected by the health services because under 45s visit the GP less often than over

45s, while that element of it that is detected may not be treated pharmacologically, as behavioural changes may suit this age group better, and most clinical evidence supporting pharmacological interventions relates to more elderly groups. As such, prescribing data will be poorer at detecting such morbidity in younger age groups. Therefore, to the extent that more elderly people are more likely to be administered pharmacotherapy to treat coronary heart disease, chronic disease scores will be better measure of disease prevalence for this group than for younger people.

For the purposes of our study, we are only interested in measuring coronary heart disease that is treated pharmacologically, since it is the only type that will affect prescribing expenditure. Therefore, we expect prevalence based on chronic disease scores to be lower than prevalence based on epidemiological evidence. Moreover, the entry criteria we use for identification of a chronic illness – at least one item prescribed per three month period over nine months – is quite strict. As a result, our estimates of chronic illness may be lower than epidemiologically derived estimates.

Table 3.13 describes psychiatric illness prevalence rates in various countries by age and gender.

Table 3.13

Prevalence of Psychiatric Illness in Ireland compared with Number with GHQ score of 4 or more in England and Scotland[1]

Group	16-24	25-34	35-44	45-54	55-64	65-74	16-74
MEN							
Ireland	*2*	*7*	*8*	*9*	*9*	*7*	*7*
Scotland	9	12	13	12	20	11	13
England	10	12	14	13	13	11	12
Northern England	10	13	17	16	14	15	14
WOMEN							
Ireland	*1*	*5*	*9*	*11*	*12*	*11*	*9*
Scotland	17	18	19	20	19	15	18
England	22	18	20	19	14	15	18
Northern England	20	20	23	20	18	17	20

Source: GMS (Payments) Board, 1999-2001; Shaw et al., 2000; Erens and Primatesta, 1999.
NOTE 1: Data in UK publications reported in integers for this table

Seven percent of men aged between 16 and 74 have a psychiatric illness in Ireland according to the chronic disease scores. This compares with prevalence estimates of between 12% and 14% for the UK studies. Nine percent of women in this age group

have a psychiatric illness in Ireland compared with between 18% and 20% in the UK studies. The prevalence of psychiatric illness in men is much more stable across age groups compared to prevalence of coronary heart disease. A stronger relationship between psychiatric illness and age is observed in Irish women than Irish men. In those aged over 45, the difference between Irish and most UK estimates of psychiatric illness prevalence differ by between 4% and 7% of the total population for men and between 4% and 9% of the total population for women. The difference is greater, however, in those aged 16 to 44. Perhaps fewer young people with psychiatric illness are treated pharmacologically. Given that it is well accepted that the level of psychiatric illness in the community is underdiagnosed and undertreated, we expected lower prevalence estimates based on the chronic disease scores than on the UK studies' questionnaire approach. These results suggest that psychiatric illness remains underdiagnosed and treated in primary care in Ireland.

In conclusion, the age-related patterns of prevalence of each illness are similar using chronic disease scores or epidemiological information. For instance, cardio-vascular disease has a strong association with age, while psychiatric illness has less of an association. For use as a risk-adjuster, we only wish to identify that proportion of a chronic illness that is treated by pharmacotherapy. As such, the lower prevalence of chronic illness measured by chronic disease scores is not particularly worrying. However, it appears that chronic disease scores cannot be used as an alternative to epidemiological evidence to measure total community morbidity as it routinely underestimates prevalence.

3.4 MISSING VALUES

In this section, we describe the extent of the missing values problem in our dataset; discuss mechanisms by which data can be missing; describe potential remedies and examine the effect of each of these remedies on the Demographic model. It represents an innovative approach to dealing with missing values in applied economics.

3.4.1 Extent of Missingness[10]

For each observation with missing values, the number of observations and the number missing as a percentage of the fully coded sample are outlined in the table below.

Table 3.14

Variables with missing values and extent of missingness

Variable	Obs	% Missing
Annual prescribing expenditure	400,751	0.00
Age	400,722	0.01
Marital	400,670	0.02
Disabil	342,689	14.49
Lonepare	392,822	1.98
GPage	394,679	1.93
Nurse	394,679	1.52
Sec	394,679	1.52
RPA	394,679	1.52
Panel	394,679	1.52
Specific	394,675	1.52
Symptomatic	394,675	1.52
Presum	394,675	1.52
Diffpan	394,679	1.52

There are 411,682 observations in the full dataset. Missingness in the response variable, because of missingness of issue date reduces the sample to 400,751. We can see in the table that the worst coded variable is disability. Almost 14.5% of observations do not have their disability coded. The extent of missingness for other variables is low, varying from only 0.01% for age to 1.98% for lone parenthood. We also find that indicators of chronic disease are always recorded when the response variable is coded, as are gender, distance, Health Board of residence, rural residence and number of people in the household. Therefore, these variables are absent from the table. Finally, there are approximately 6,000 patients whose GP-level variables have missing values. However, these patients either do not have a chronic disease recorded as they were not on the GMS scheme during the chronic disease measurement period, or they attend a GP whose panel size varied by more than 10% from December 2000 to October 2001 and thus were excluded as 'unusual' GPs. Therefore, those with missing GP values are dropped automatically for the Supply model (which is the only model that uses GP variables) and we do not need to multiply impute their values.

[10] An ugly word, but standard in the literature.

Since the disability variable has the greatest level of missingness, it is worth further exploratory investigation. Table 3.15 illustrates the differences in average effects of a number of variables for people with their disability field coded versus people whose disability field is missing.

Table 3.15

Average Effects for Selected Variables depending on whether Disability is Coded or Missing

Variable	Average when disability is coded	Average when disability is missing
Totcost (IR£)	221.68	150.95
Age	45.00	34.33
Gender	0.54	0.67
Numcard	2.31	3.15
NEHB	0.30	0.03
SEHB	0.22	0.86
SHB	0.48	0.11

Average annual prescribing expenditure is IR£221.68 for those whose disability field is coded, whereas it is only IR£150.95 when disability is missing. This is probably due in large part to the fact that younger people are more likely to have their disability field missing – the average age for those with a missing disability field is 34.33 versus 45 for those with disability coded. Those with disability missing are more likely to be female, come from larger families and, most notably, come from the South Eastern Health Board. We find that 86% of all those whose disability field is missing come from the South Eastern Health Board.

The effect of missingness differs by model, as outlined in Table 3.16.

Table 3.16

Sample Size, Complete Cases and Missing Cases by Model

Model	Sample Size	Complete Cases	Missing Cases
NARA	400,751	400,722	29
Demographic	400,751	342,483	58,268
Chronic Illness	362,072	310,892	51,180
Supply	305,887	261,513	44,374

The effect of missingness on the NARA is tiny – only 0.01% of observations have missing values. The incremental effect of missingness for the Demographic is considerable, however. Over 14.5% of observations have missing values, reducing

sample size from 400,751 to 342,483. There is no increased missingness in the Chronic Illness model as chronic illness indicators are fully coded. However the inclusion criteria for the Chronic Illness model are stricter, because individuals had to be present in the chronic illness measurement period as well as during the cost year for inclusion in the Chronic Illness model. If fully coded there would be 362,072 people in the Chronic Illness model. Missingness in the demographic and socio-economic variables reduces this to 310,892 (there is a greater number of missing values in the Demographic model than in the Chronic Illness model because some of those observations with missing values in the Demographic variables do not meet the inclusion criteria of the Chronic Illness model).

Finally, the inclusion criteria for the Supply model are even stricter. As discussed above, 169 GPs out of 625 had their panels rise or fall by more than 10% from December 2000 to October 2001. In addition, 1.52% of patients report GP identifiers did not match any GP in our dataset. Since we have no information about these GPs, we cannot sensibly impute missing data for them (except to do so from their patient characteristics, which is probably to generate spurious relationships). Therefore, the study sample is 305,887. Missing values reduces this to 261,513.

3.4.2 Mechanisms of Missing Data

The mechanism by which data are missing can be of three types, depending on our ability to predict missing values – missing completely at random, missing at random and nonignorable. The terminology of King et al. (2001) is used to describe these three concepts. Let D denote the data matrix, including the response variable Y and covariates X, such that $D = \{Y, X\}$. Let M be a matrix indicating missing values with the same dimensions as D, where in M a 1 is inserted for a coded record and a 0 for a missing value. Let D_{obs} and D_{mis} represent portions of D for which D is observed and missing respectively, such that $D = \{D_{obs}, D_{mis}\}$. Missing completely at random (MCAR) are missing values that cannot be predicted using the dataset, or M is independent of D: $P(M/D) = P(M)$. It is rare that data are MCAR. Missing at random (MAR) means that a missing value can be predicted from D_{obs}, but must be independent of D_{mis} having controlled for D_{obs}, that is, assuming that M is independent of D_{mis}: $P(M/D) = P(M/D_{obs})$. Finally, missing values can be nonignorable (NI), which is where the probability of being missing depends on the

value that is missing and not on other variables in the dataset. An example of NI is when high income respondents to a survey are less likely to code their income, and no other question in the survey predicts if someone is high income.

Most empirical analysis of missing data finds that data are MAR rather than MCAR. Meanwhile, we cannot prove NI using observed data only since whether or not missing data are NI cannot be predicted from available data and we cannot gauge the importance of the missing data because they are missing! Therefore, we cannot choose empirically between MAR and NI. In this study we will show below that the probability of the disability field being missing appears to be related to prescribing expenditure, residence in the South Eastern Health Board, age, gender and number of people in the household with a medical card, suggesting that the pattern of missing data is more likely to be MAR than MCAR or NI.

3.4.2.1 Approaches to Dealing with Missing Values

The three principal methods for data analysis in the presence of missing data are listwise deletion, least squares imputation and multiple imputation.

First, listwise deletion considers only those observations that are coded in all fields that are being analysed. This is the default approach of most commercial statistical software packages. However it discards much useful information in the dataset and is biased unless MCAR holds (King et al. 2001). Nevertheless, it is the default option of most data analysis in social science[11].

Second, least squares methods impute missing values in one of two ways. The unconditional mean of each variable for which there are missing values can simply replace any missing values. Alternatively, dummy variables indicating missing values for each variable are generated, each one is regressed on all other regressors and predicted values inserted in place of the missing values. Weighted least squares should be used, to downweight the effect of imputed values, where the weight applied is the ratio of the residual variance of Y given all the X's to the residual variance of Y given the observed X's for each case (Little, 1992: 1231).

[11] King et al. (2001:49) found that *"almost all"* political science analysts fill in some items with educated guesses and 94% used listwise deletion for any remaining missing values.

Third, multiple imputation involves imputing m values for each missing values and creating m datasets, in which the observed data do not change but the imputed values are different due to the stochastic process of imputation. Results are generated for each of the datasets and combined afterwards. Unlike least squares imputation, multiple imputation introduces random error into the imputation process, making it possible to get approximately unbiased estimates of all parameters, so long as data are MAR (Allison, 2000). Moreover, the standard errors about results generated by multiple imputation are usually greater than those generated by methods such as listwise deletion, as they reflect the uncertainty due to imputation. Because of these advantages and its general applicability, multiple imputation is becoming the recommended approach to handling missing data in social sciences (Rubin, 1996; King et al., 2001)

We borrow heavily from King et al. (2001) to describe the following multiple imputation model. Variables are assumed to be jointly multivariate normal, an assumption that appears to work as well as more complex alternatives. For observation i (i = 1,…,n), let D_i denote the vector of values of all variables with mean vector μ and variance matrix Σ. The likelihood function for complete data is

$$L(\mu, \Sigma \mid D) \propto \prod_{i=1}^{n} N(D_i \mid \mu, \Sigma). \tag{3.6}$$

Let $D_{i,obs}$ denote the observed elements of row i of D, and $\mu_{i,obs}$ and $\Sigma_{i,obs}$ denote the corresponding sub-vector and sub-matrix of μ and Σ. These do not vary over i. The likelihood function for observed data is:

$$L(\mu, \Sigma \mid D_{obs}) \propto \prod_{i=1}^{n} N(D_{i,obs} \mid \mu_{i,obs}, \Sigma_{i,obs}). \tag{3.7}$$

Given the multi-variate normal specification, missing values are imputed linearly, analogous to simulations from a regression. For instance, let $\widetilde{D}_{ij}$ denote an imputed value for observation i and variable j and let $D_{i,-j}$ denote the vector of values of all observed variables for observation i except variable j. Taking random draws from the posterior distribution of μ and Σ, we regress D_j on D_{-j}, as follows:

$$\widetilde{D}_{ij} = D_{i,-j}\widetilde{\beta} + \widetilde{\varepsilon}_i \tag{3.8}$$

where ~ represents a random draw and ε_i is the error term.

The difficulty in this approach is taking random draws from the posterior of μ and Σ. Two commonly used approaches include the Imputation Posterior (IP) and Expectation Maximisation (EM) algorithms. The IP algorithm is slow and because it is based on Markov Chain Monte Carlo methods, judging convergence requires considerable expertise. On the other hand EM is fast, as it uses a deterministic algorithm to find the maximum of the likelihood function, but it does not yield the rest of the distribution. Therefore we use EMs, which is EM with sampling.

EM essentially makes one-step maximum likelihood easier by using a sequence of simpler maximisations. Each iteration of the EM algorithm consists of two steps – the E step and the M step. In simple cases such as in this study, EM involves running regressions to estimate β, assigning predicted values to missing values (the E step), re-estimating β (the M step), and iterating until convergence. The E step and M steps can be expressed more formally in equations (3.9) and (3.10) respectively:

$$Q(\theta / \hat{\theta}^{old}) = E[D_i / \hat{\theta}^{old}] \qquad (3.9)$$

$$\hat{\theta}^{new} = \arg\max_{\hat{\theta} \geq 0} Q(\theta / \hat{\theta}^{old}) \qquad (3.10)$$

where $\hat{\theta} = vec(\hat{\mu}, \hat{\Sigma})$, where the vec(.) operator stacks the unique elements.

Because EM does not yield the whole distribution, uncertainty due to imputation is ignored (akin to least squares treatment of missing values), leading to underestimated standard errors. However, EMs overcomes this drawback by beginning with EM and then adding back uncertainty due to imputation using a Bayesian method known as 'data augmentation'. Having run EM, we find the maximum posterior estimates of the parameters $\hat{\theta}$. Then we compute the variance matrix, $V(\hat{\theta})$. Next we draw a simulated θ from the normal distribution with mean $\hat{\theta}$ and variance $V(\hat{\theta})$. Next we compute $\tilde{\beta}$ deterministically, simulate $\hat{\varepsilon}$ from the normal distribution and substitute these values into Equation (3.8) above to generate an imputed value. This process is repeated m times to create the required number of datasets. According to King et al. (2001:55) *"EMs is very fast, produces independent imputations, converges nonstochastically, and works well in large samples"*.

Five imputed values are usually considered sufficient to reflect imputation uncertainty. The parameters of the five imputed datasets are combined such that the results reflect imputation uncertainty as well as model uncertainty. The point estimate of any parameter, q^*, is:

$$q^* = \frac{1}{m}\sum_{j=1}^{m} q_j , \tag{3.11}$$

where m is the number of imputations and j is an imputed value. The standard error for any parameter incorporates within dataset variance and variance between imputed datasets as follows:

$$SE(q)^2 = \frac{1}{m}\sum_{j=1}^{m} SE(q_j)^2 + S_q^2(1+\frac{1}{m}), \text{ where } S_q^2 = \sum_{j=1}^{m}\frac{(q_j - q^*)}{(m-1)}. \tag{3.12}$$

Multiple imputation is conducted in this study using Amelia 2.0 (Honaker et al., 1999), while results are combined in Stata 7.0 (StataCorp, 2000) using the Clarify suite of programmes (Tomz et al., 2001).

3.4.3 Multiple Imputation of Missing values: Results

Table 3.17 illustrates the effect of missingness on efficiency and bias in the Demographic model.

Table 3.17

List-wise Deletion versus Multiple Imputation for Demographic Model[1]

	List-wise Deletion		Multiple Imputation	
Variable	Co-effic.	P>|t|	Co-effic.	P>|t|
Age	4.33	0.00	4.55	0.00
Agesq	0.00	0.05	0.00	0.61
Gender	-1.58	0.30	-5.94	0.00
Marital	27.84	0.00	25.55	0.00
Disabil	147.31	0.00	127.92	0.00
Lonepare	-5.04	0.07	-2.50	0.46
Rural	-11.07	0.00	-12.79	0.00
Numcard	-19.32	0.00	-19.77	0.00
3-5 miles	-10.11	0.00	-9.53	0.00
5-7 miles	-11.11	0.00	-10.98	0.00
7-10 miles	-32.66	0.00	-32.31	0.00
10+ miles	-25.86	0.00	-25.01	0.01
NEHB	15.21	0.00	10.83	0.06
SEHB	27.24	0.00	11.53	0.05
_cons	40.41	0.00	42.35	0.00
Root MSE		435.98		425.52
N		342,483		400,751

NOTE 1: Relates to all prescribing expenditure, unlike later models where it is truncated and budget neutral drugs are removed. See Table A5.2 in Appendix 5.1 for an examination of the effects of truncation.

The efficiency of the multiply imputed model appears to be better than that of the list-wise deletion model. First, it uses 400,751 observations as against 342,483 observations for the list-wise deletion model. Second, it has a lower Root Mean Squared Error, meaning that the explained variance of the multiply imputed model exceeds that of the list-wise deletion model.

Coefficients differ between the two models as well. The four variables whose coefficients change the most are disability, SEHB, NEHB and gender, in rank order of magnitude of effect. Bivariate analysis above showed that those for whom disability was not coded had lower average expenditures than those with disability coded. Therefore including these people in the analysis reduced the value of the disability coefficient by IR£18.06. Consequently, the value of the coefficient on SEHB, the health board with by far the greatest number of missing disabled, also drops considerably.

3.4.4 Conclusions

As with the majority of public use datasets, and especially administrative datasets, there is a considerable proportion of missing values in our dataset. These are concentrated in the disability variable, but variables such as lone parenthood, marital status and some GP-related variables suffer from missingness as well. The conventional method of data analysis in the presence of missing data is list-wise deletion. We find that such an approach would reduce sample sizes by up to 15%. Alternatives to list-wise deletion include least squares and multiple imputation. Multiple imputation offers a flexible and statistically valid approach to accounting for missing data. We find that applying multiple imputation to the Demographic model leads to a greater level of explained variance and changes in coefficient values when compared to list-wise deletion. These changes are in line with exploratory bivariate analysis of missingness in the disability field.

In a review of the use of multiple imputation and bootstrapping in econometrics, Brownstone and Valetta (2001) mention only one specific application of multiple imputation (Brownstone and Valetta, 1996). It has clear advantages over alternative approaches, is easy to apply and is likely to have an increasing influence on handling missingness in future empirical analysis in economics.

3.5 COLLINEARITY OF COVARIATES

There is no evidence of serious collinearity in the covariates. The greatest partial correlation coefficients are of the order of 0.25 to 0.28. Age is positively correlated with marriage, rural status and distance from GP and negatively correlated with lone parenthood. Gender is positively correlated with lone parenthood and negatively related to being married, presumably due to the high number of widows in the elderly population. Disability is negatively related to lone parenthood, possibly due to an artefact of the data. Lone parenthood is negatively related to marriage. Distance is positively related to rurality and age.

Of the chronic disease indicators, coronary heart disease has the strongest positive bi-variate relationship with other indicators. This may well be because those on treatment for coronary heart disease often have co-morbidities such as diabetes. Other indicators with noticeable correlations include rheumatological conditions, pain and psychiatric illness.

3.6 DATA FOR MODEL EVALUATION

The preceding five sections were concerned with the data used to estimate models. The goal of risk-adjustment is to ensure unbiased and efficient predictions of future health care utilisation, as well as consider the distributional implications of those predictions. This section considers variables that can be used to assess the distributional implications of competing models of risk-adjustment. Chief of these is income. Section 3.6.1 describes a method for imputing income for all individuals on the Medical Card Register, which is another significant contribution in this chapter. Section 3.6.2 describes other variables used for model evaluation.

3.6.1 Imputing Income for Medical Card Register

Although eligibility for GMS services is means-tested, income is not included on the Medical Card Register. However a number of predictors of income are included and we exploit similarities between the Medical Card Register and the Household Budget Survey 1999/2000, which contains an income variable, to impute income for all individuals in our sample.

3.6.1.1 *Imputation: Data and Methods*

The Household Budget Survey 1999-2000 (2002), undertaken between June 1999 and July 2000 was a nationally representative survey of income and household expenditure of 7,644 households, of whom 3,019 (39%) were households in receipt of GMS. There are a number of household characteristics common to both the Household Budget Survey and the Medical Card Register, namely age, gender, marital status, lone parenthood and morbidity. Micro-data on these variables as well as weekly income was gratefully received from the CSO[12]. We wish to impute income for everyone on the Medical Card Register using data from the Household Budget Survey.

The most obvious approach to imputing income for households on the Medical Card Register based on the Household Budget Survey (2002) would be to treat income in the Medical Card Register as a missing data problem, and estimate it using the multiple imputation techniques described above. It is not possible to do this directly however, because of the absence of information on the two types of household on the medical card scheme – those households where everyone is on the scheme (full GMS household), and those households where only a proportion is on the scheme (partial GMS household). Of the 3,019 households in receipt of GMS in the Household Budget Survey, 2,247 are full GMS households and 772 are partial GMS households. Disposable income of the latter household type is IR£483.52 on average, while it is IR£192.20 on average for the former household type. Therefore, these are very different types of household with respect to disposable income and we wish to distinguish between them before predicting income in the Medical Card Register.

While there is no indication in the Medical Card Register whether a person is in a full GMS household or a partial GMS household, we can predict the probability of a household being a member of each group using a probit model, as follows:

$$\text{Prob}[G_i = 1] = \Phi(\alpha x_i), \qquad (3.13)$$

where G_i is one if the household is a full GMS household and zero otherwise, x_i is a row vector of relevant household characteristics, α is the set of parameters to be estimated and $\Phi(.)$ is the standard normal distribution.

[12] The data were received in advance of the CSO's intended date of distribution of micro-data, for which we are grateful.

We then estimate the relationship between household characteristics and income for each type of household as follows:

$$(I_i \mid G_i=0) = x_i\beta + e_i \tag{3.14}$$

$$(I_i \mid G_i=1) = x_i\gamma + v_i \tag{3.15}$$

where I_i is the income of household i, β is the set of parameters relating to partial GMS households, γ is the set of parameters relating to full GMS households, and e_i and v_i are i.i.d error terms. Therefore, predicted income for households in the Medical Card Register is derived as:

$$\hat{I}_i = (p_i)(\hat{I}_i / G_i = 1) + (1 - p_i)(\hat{I}_i / G_i = 0), \tag{3.16}$$

where p_i is the probability of household i being a full GMS household and $\hat{I}$ is predicted income.

One adjustment to this model is made for practical purposes. In the Medical Card Register, 3,064 people are identified as being under 14 and being the only person in their family with a medical card. Rather than rely on the estimates of the probit model, these get a probability of membership of the partial GMS group of one.

Owing to the positive skew that the data usually exhibit, the income variable is often log-transformed. We examined both a linear and log-linear specification and chose the linear one because its explained variance was as high and it is easier to interpret. The results of the log-linear specification are presented in Appendix 3.4.

Although we cannot perform multiple imputation directly on the Medical Card Register, we could impute multiple observations for full GMS households and partial GMS household, equivalent to equations (3.14) and (3.15) above, and then multiply each of these observations by the probability of membership of each group for each household, as per equation (3.13). Since we have already produced multiple observations for those missing values that we imputed in the Medical Card Register, to produce another set of multiple observations would produce an extremely cumbersome model. In addition, income data are used in validation rather than estimation and are needed to rank individuals, so the data need only be ordinal rather

than cardinal. Therefore, a simpler imputation technique than multiple imputation is adequate.

3.6.1.2 Equivalence Scales

Prescribing analysis is done at individual rather than household level. Therefore, we need to equivalise household income to express it relative to the income of a representative individual. An equivalence scale is an attempt to account for differences in purchasing power of different sized households, reflecting the fact that larger households enjoy economies of scale through the joint consumption of certain goods, such as housing, and the fact that children do not consume as much as adults.

A large literature exists on the appropriate means of equivalising household income. Coulter et al., (1992) distinguish between five estimation methods for equivalence scales. First, there are scales based econometric analysis of household consumption patterns, which use approaches based on Engel (1895) or the Barten method that determine the additional consumption of each additional household member and construct equivalence scales on this basis. Second, there are subjective scales, based on the opinions expressed by consumers. Third, there are the implicit scales applied by government services to calculate household's eligibility for various means-tested public services. Fourth, there are equivalence scales estimated from general household characteristics. Finally there are scales based on expert opinion, the most commonly used one being the OECD scale which applies a weight of 1 to the first adult in the household; 0.7 to each additional adult and 0.5 to each additional child.

Cutler and Katz's (1992) approach is to equivalise income based on the expression $E = (A+cK)^e$ where E is equivalent income, A is number of adults in the household, K is number of children, c is the cost of a child relative to an adult and e is the economies of scale associated with larger households. They apply values of e=0.5 and c=0.4.

Jesuit and Smeeding (2002) used an equivalence scale of e=0.5 and c=1 in an analysis of international income inequality. Citro and Michael (1995) recommend that e vary from 0.65 to 0.75 and c vary from 0.7 to 1. The OECD scale is commonly used in international comparisons of inequality and poverty including

many applications of the Luxembourg Income Study, such as Rostek (2000). In Ireland, Nolan and Russell (2001) used an implicit social welfare scale of 1:0.66:0.33 for baseline calculations, and applied the OECD scale, as well as a 1:0.6:0.4 scale as part of their sensitivity analysis.

Cowell and Mercader-Prats (1999) recommend that a variety of equivalence scales be used as one of the ways of mitigating the effects of choice of equivalence scale on ultimate results. Therefore, for baseline calculations we apply an e of 0.5 (square root), similar to Cutler and Katz (1992) and a c of 0.7, similar to the OECD scale and Citro and Michael's (1995) recommendations. This produces values similar to Nolan and Russell's (2001) baseline equivalence scale for Ireland for most household sizes. For comparison, we apply the old OECD scale, as it is popular for international inequality comparisons.

3.6.1.3 Results

Table 3.18 describes the distribution of income in the Household Budget Survey.

Table 3.18

Distribution of Household Income (IR£)

Statistic	Full GMS	Partial GMS	All GMS	All HBS[1]
Mean	192.20	483.53	266.69	417.58
Median	159.98	438.28	191.79	359.33
25th	99.99	313.83	118.87	190.06
75th	234.87	600.02	343.48	570.58
N	2,247	772	3,019	7,644

Source: Household Budget Survey 1999/2000, 2002.
NOTE 1: HBS = Household Budget Survey

For all households surveyed, average weekly disposable income was IR£417.58, while for households with GMS recipients it was IR£266.69 or 63% of average income. For those where all members were GMS recipients income was only IR£192.20 or 46% of average income, whereas income for those where some but not all members of the household were GMS recipients was IR£483.53 or 116% of average income.

The data are positively skewed, as with most incomes data, so medians are less than means for all sub-groups and in total. The inter-quartile range for households where

all are GMS recipients is IR£99.99 to 234.87 (range is -IR£1.66 to IR£1,732.50).[13] Meanwhile the inter-quartile range for households where some but not all members are GMS recipients was much higher, higher indeed than the inter-quartile range for all respondents.

Table 3.19 describes those variables that the Household Budget Survey and the Medical Card Register have in common.

Table 3.19

Description of Variables common to both Household Budget Survey and Medical Card Register

Variable	Description
%m00_05	Percentage in the household male aged 0-5
%m06_13	Percentage in the household male aged 6-13
%m14_20	Percentage in the household male aged 14-21
%m21_44	Percentage in the household male aged 21-44
%m45_64	Percentage in the household male aged 45-64
%m65etc	Percentage in the household female aged 65 or over
%f00_05	Percentage in the household female aged 0-5
%f06_13	Percentage in the household female aged 6-13
%f14_20	Percentage in the household female aged 14-21
%f21_44	Percentage in the household female aged 21-44
%f45_64	Percentage in the household female aged 45-64
%f65etc	Percentage in the household female aged 65 or over
Marital	Household headed by a married person
Incaporill	Incapacitaty or Ill
Lonepare	Household is headed by a lone parent

Data are available for the household head, principal household member and chief economic supporter in each household from the Household Budget Survey. These are the same person in many households. In the Household Budget Survey the incapacity variable refers to an incapacity or illness by any of the above three named persons. The equivalent variable in the Medical Card Register is whether a household is in receipt of disability payments. Most people who are in receipt of these payments would state in the Household Budget Survey that they are incapacitated or unable to work due to illness or injury. Therefore, the variables are reasonably well matched, but it is possible that people outside these three are in receipt of disability payments. In this case, they will not be measured by the

[13] Negative incomes are possible, for instance in the situation where a household paid a tax bill in the previous week that exceeded its gross income in the previous week

incaporill variable in the Household Budget Survey and will be measured in the Medical Card Register.

Table 3.20 compares the prevalence of each of the variables that the Household Budget Survey and the Medical Card Register have in common.

Table 3.20

Descriptive Statistics of Variables Common to Household Budget Survey and Medical Card Register

Variable	HBS*	MCR*	MCR/HBS
m00_05	0.03	0.04	1.19
m06_13	0.07	0.06	0.81
m14_20	0.07	0.05	0.72
m21_44	0.11	0.09	0.84
m45_64	0.10	0.10	0.99
m65etc	0.09	0.11	1.26
f00_05	0.03	0.04	1.10
f06_13	0.08	0.06	0.74
f14_20	0.06	0.06	0.88
f21_44	0.13	0.13	0.95
f45_64	0.11	0.11	1.02
f65etc	0.12	0.17	1.42
Marital	0.51	0.31	0.61
Incaporill	0.12	0.16	1.31
Lonepare	0.05	0.07	1.39
N	3,019	248,080	

*HBS = households with at least some GMS recipients in the Household Budget Survey 1999/2000; MCR = Medical Card Register 2000.

The table describes the proportion of the average household in each age and gender category as well as the proportion of households headed by married persons, incapacitated or ill people and lone parents, according to the Household Budget Survey and the Medical Card Register. Note that the 400,751 individuals in the Medical Card Register form 248,080 households.

The second and third columns give the prevalence rate of each variable in the Household Budget Survey and Medical Card Register respectively, while the fourth column compares each rate. We expect the comparisons to be reasonably similar, since they are taken from similar populations at similar points in time and find that they are so. They differ greatest in the proportion of females aged over 65, marital status, incapacitated or ill and lone parenthood. One explanation why the Medical Card Register reports a greater proportion of females aged over 65 than does the

Household Budget Survey is that the Household Budget Survey does not include people living in long stay institutions such as nursing homes. The difference in lone parenthood variable may be a sampling issue in the Household Budget Survey, since in 1995 Household Budget survey reported prevalence of 9% for this variable, as described in Table 3.9, compared with 5% above. The difference in the incaporill variable may be due to definitional differences as explained above.

Table 3.21 provides comparative statistics for households with full GMS eligibility and households with partial GMS eligibility.

Table 3.21

Comparison of Variables between Households with Full GMS Eligibility and Households with Partial GMS Eligibility in Household Budget Survey

Variable	FullGMS*	PartialGMS*	pGMS/fGMS*
M00_05	0.04	0.02	0.65
M06_13	0.08	0.06	0.75
M14_20	0.06	0.08	1.39
M21_44	0.08	0.16	2.08
M45_64	0.09	0.12	1.32
M65etc	0.11	0.05	0.48
F00_05	0.04	0.02	0.67
F06_13	0.09	0.05	0.60
F14_20	0.06	0.07	1.26
F21_44	0.12	0.15	1.25
F45_64	0.10	0.13	1.27
F65etc	0.14	0.07	0.51
Marital	0.47	0.65	1.38
Incaporill	0.13	0.10	0.74
Lonepare	0.07	0.00	0.06
N	2,247	772	

* fullGMS (fGMS) = households where all members are GMS recipients in the Household Budget Survey 1999/2000; partialGMS (pGMS) = households where some but not all members are GMS recipients in the Household Budget Survey 1999/2000.
Source: Household Budget Survey 1999/2000 (2002)

We expect the households where all members are GMS recipients – the full GMS households – to differ from those where some but not all members are GMS recipients – the partial GMS households - with respect to the variables described. We find that there is a higher proportion of males in economically active age groups (14-64) and less males aged over 65 in those households where some but not all are GMS recipients. A similar pattern exists for females but the difference is not as great. There are more households headed by married persons, fewer households with

incapacitated or ill members and far less lone parents in the sample of partial GMS households. These findings are as expected.

Following on from Table 3.21, Table 3.22 presents the results of a probit analysis of membership of families where all members are GMS recipients.

Table 3.22

Factors Affecting Membership of Households where all members are GMS Recipients

Variable	Coef.	P|z|>0
M06_13	-0.002	0.968
M14_20	-0.216	0.001
M21_44	-1.347	0.000
M45_64	-1.285	0.000
M65etc	-0.894	0.000
F00_05	-0.070	0.505
F06_13	0.145	0.020
F14_20	-0.211	0.000
F21_44	-1.107	0.000
F45_64	-1.154	0.000
F65etc	-0.931	0.000
Marital	0.584	0.000
Incaporill	0.513	0.000
Lonepare	0.464	0.069
_cons	2.613	0.000
Pseudo-R^2	0.3367	
Prob > Chi2	0	
N	3,019	

*Heteroscedasticity robust standard errors

Source: Household Budget Survey, 1999/2000 (2002)

The reference class is males aged 0 to 5. Most other age groups are less likely to be in households where all members are GMS recipients than the reference class, with the exception of females aged 6 to 13, while males aged 6 to 13 and females aged less than 6 are insignificantly different from the reference class. Households headed by married persons as well as households with incapacitated or ill members are more likely to be from full GMS households, while households headed by lone parents are insignificantly different from the reference class.

Table 3.23 describes the factors affecting household income for those households where all members are GMS recipients.

Table 3.23

Factors affecting Household Income for those Households where all Members are GMS Recipients

Variable	Coef.	P|t|>0
M06_13	17.19	0.00
M14_20	46.40	0.00
M21_44	105.62	0.00
M45_64	88.77	0.00
M65etc	82.70	0.00
F00_05	22.26	0.01
F06_13	11.53	0.05
F14_20	54.23	0.00
F21_44	119.42	0.00
F45_64	94.37	0.00
F65etc	74.01	0.00
Marital	-6.02	0.46
Incaporill	-36.62	0.00
Lonepare	18.21	0.05
_cons	27.24	0.01
R^2	0.4400	
Prob >F	0.0000	
N	2,247	

*Heteroscedasticity robust standard errors

Source: Household Budget Survey, 1999/2000 (2002)

The model is significant and the R^2 of 0.44 is high for a cross-sectional regression. The reference category is again males aged 0 to 5. The main factors affecting household income are the number of people in aged 21 to 64, especially males aged 21 to 64. Lone parents families in this group are insignificantly different from the reference class, while households with incapacitated or ill members have lower than average incomes. Marital status has a negative sign, but it is insignificant.

Table 3.24 describes the factors affecting income for households where some but not all members are GMS recipients.

Table 3.24

Factors affecting Income for Households where Some but not all Members are GMS Recipients.

Variable	Coef.	P|t|>0
M06_13	26.26	0.06
M14_20	67.13	0.00
M21_44	179.96	0.00
M45_64	153.66	0.00
M65etc	106.78	0.00
F00_05	54.12	0.02
F06_13	22.83	0.14
F14_20	73.27	0.00
F21_44	161.98	0.00
F45_64	143.24	0.00
F65etc	132.11	0.00
Marital	-20.07	0.23
Incaporill	-85.90	0.00
Lonepare	-23.37	0.58
_cons	35.06	0.22
R^2	0.4087	
Prob >F	0.0000	
N	772	

*Heteroscedasticity robust standard errors

Source: Household Budget Survey, 1999/2000 (2002)

The model is again significant and has a high level of explained variance. The reference case is again males 0 to 5. As with the last model, the number of members aged 21 to 64 has a large effect on household income, although in this instance persons aged over 65 is also important. Lone parenthood and marital status are insignificantly related to household income and number of household members with an incapacity or illness is negatively related to income.

Table 3.25 describes the distribution of predicted equivalent income for the Medical Card Register population, according to the baseline equivalence scale and old OECD scale as described above. For comparison, equivalent income for the Household Budget Survey respondents who are from households with at least some GMS recipients is also presented.

Table 3.25

Predicted Equivalent Income for Medical Card Register and Equivalent Income for Household Budget Survey[14]

Group	Mean	Median	25th	75th	N
MCR Base	128.54	130.39	104.30	149.32	400,751
MCR OECD	106.45	105.12	94.15	120.19	400,751
HBS Base	159.04	126.40	95.09	190.84	3,019
HBS OECD	120.68	100.09	83.02	140.59	3,019

Source: Household Budget Survey, 1999/2000 (2002); Medical Card Register, 2000

The average equivalent income in the Medical Card Register is IR£128.54 according to the baseline equivalence scale and IR£106.45 according to the old OECD scale. By comparison, the average equivalent income in the Household Budget Survey for households with at least some GMS recipients was IR£159.04 according to the baseline scale and IR£120.68 according to the old OECD scale. This difference in mean incomes is discussed below.

As is normally the case with predicted values, the variation is less for predicted equivalent income than for actual equivalent income. The inter-quartile range for the baseline prediction in the Medical Card Register is IR£104.30 to IR£149.32 whereas for the baseline equivalent income in the Household Budget Survey the inter-quartile range is much larger at IR£95.09 to IR£190.84.

3.6.1.4 Discussion

We wish to assess the distributional consequences of various formulae for prescribing budget setting with respect to policy relevant variables. Poverty is one such variable, which is usually measured as income poverty. Although household income is not collected as part of the Medical Card Register, it is collected as part of the Household Budget Survey. Consequently, we have been able to impute household income for each individual on the Medical Card Register based on the set of relevant variables that are collected in both the Medical Card Register and Household Budget Survey.

[14] Since there are some children who are medical cardholders in the Medical Card Register, as described above, the equivalence scale for these is 0.84 [= $(.7)^{0.5}$] in the baseline and 0.5 in the case of the old OECD scale, not 1 as is commonly the case.

We find that the set of variables common to both datasets explains a large degree of variation in household income in the Household Budget Survey. As such, it is a good set of predictors. The two datasets differ insofar as households where some but not all members are GMS recipients are identified in the Household Budget Survey but not in the Medical Card Register. We find, however, that whether or not your household is exclusively GMS or not depends on the number of people of working age in the household as well as marital status and number with an incapacity or illness in the household. As such we can predict membership of each group and impute income based on probability of membership of each group and predicted income for each group.

If the Household Budget Survey was perfectly representative of the GMS population, and if all variables in both this survey and the Medical Card Register were measured in the same way and without error, and if our model predicting income was well-specified, then we would expect the mean and distribution of equivalent income to be very similar in both datasets. We find that mean equivalent income is 19% lower in the Medical Card Register than in the Household Budget Survey. There are two points worth noting.

First, while mean incomes differ, we are more interested in income rankings for distribution analysis. Median incomes are almost identical and income at the 25th percentile according to the Medical Card Register actually exceeds that according to the Household Budget Survey.

Second, the difference in means may be due to a difference in the measurement of disability. One of the principal negative associations with household income in the Household Budget Survey was found to be the number of people in the household who were permanently incapacitated or who were unable to work due to illness or injury. Meanwhile, the equivalent variable in the Medical Card Register is whether a household is in receipt of disability payments. We find that there is a higher prevalence of disability in the Medical Card Register than there is of incapacity or illness in the Household Budget Survey in Table 3.19. The difference in mean incomes could be due to differences in the measurement of incapacitated or ill, leading to differences in their relative prevalence, leading to differences in income

predictions. We cannot rule out measurement error in the Medical Card Register either, since disability is a discretionary field and is therefore less likely to be coded accurately, while we found that there were a large number of missing values for this field in the South Eastern Health Board. It appears that disability is the principal reason for the difference in average income between the two datasets.

As long as measurement error and differences in the measurement of the disability variable in both the Household Budget Survey and the Medical Card Register are unrelated to household income, and we have no reason to suspect that it would be related to income, then income rankings should not be disturbed unduly. As such, we can proceed with distributional analysis based on predicted income in the Medical Card Register.

3.6.2 Other Variables Used in Model Evaluation

Additional variables relating to vulnerable groups are collected by some health boards but not universally. Therefore, while we cannot use them as variables in modelling the factors affecting prescribing expenditure, they can be used for model evaluation. We favour the model of prescribing that minimises prediction error for each group. Table 3.26 outlines each of these variables and their frequency.

Table 3.26

Description of Vulnerable Groups

Variable	Freq.
Living Alone	8,301
Chronic Illness	2,110
Financial Hardship	7,023
Drugs and Hardship	1,687
Refugee / Asylum Seeker	482
Unemployment Assistance	12,847
Supplementary Welfare Allowance	3,239
Early School Leaver	307

There were 4,029 persons indicated as living alone in the South Eastern Health Board and 4,272 in the Southern Health Board, 8,301 in total. They had an average age of 69. There were 2,110 people indicated as having chronic illnesses in the South Eastern Health Board. They had an average age of 52. It is not recorded how the Health Board defines a chronic illness, so the relationship between these data and chronic disease scores cannot be established.

Hardship cases are people who get medical cards on discretionary grounds rather than on income grounds. They often have chronic illnesses requiring regular GP consultations and many are on long-term medication. There are 4,883 such cases reported in the South Eastern Health Board and 2,140 in the Southern Health Board. They had an average age of 49. In addition, the Southern Health Board has a category called Drugs and Hardship who number 1,687. These usually got a medical card originally because they were on long term medication. Owing to their similarity, these categories are combined in the following chapters.

There are 150 refugees and asylum seekers recorded in the North Eastern Health Board and 332 in the Southern Health Board. They have an average age of 25. These are a politically sensitive group, with claims that they are high users of health services. As such GPs are likely to favour models that predict the expenditures of this group accurately.

Unemployment assistance is given to job seekers who do not have sufficient social insurance credit to qualify for unemployment benefit (which is a higher weekly payment) or who are unemployed for so long that they have used up their social insurance entitlement. It can be treated as an indicator of poverty. There are 12,847 people from households in receipt of unemployment assistance in the Southern Health Board.

Supplementary Welfare Allowance is a health board payment to households who are unable to manage financially on social welfare or other income sources, usually due to exceptional circumstances. There are 3,239 people reported as coming from households in receipt of Supplementary Welfare Allowance in the Southern Health Board.

The Southern Health Board Medical Card Register includes data on people aged 16 to 18 who have no income, are living with their parents and are not at school. We refer to them as early school leavers. There are 307 in the dataset.

3.7 CONCLUSIONS

This study has generated a unique dataset on GMS prescribing expenditure, demographic and socio-economic status, health status and access to GP services for GMS recipients, as well as data on GP characteristics. As such, we generated variables in each of the five categories of variable described in Chapter 2, namely, demographic, socio-economic, health status, access to services and physician characteristics. This chapter described and evaluated the dataset, imputing missing values and a new income variable. Six significant contributions were noted.

First, the Medical Card Register was unearthed as a source of demographic, socio-economic and access variables. Second, chronic illness indicators were constructed. Third, a test for measurement error in the chronic illness indicators was applied. Fourth, new variables measuring GP prescribing style were generated. Fifth, multiple imputation of missing records was applied. Sixth, a new income variable was imputed for all observations.

The chapter began by laying out the inclusion criteria for the study and explained the process of inflating expenditure for individuals who were GMS recipients for part of the cost year. Steps taken to harmonise datasets were described. Section 3.2 was a long section describing the variables used in the study and assessing their reliability and validity. It found that the number of errors in the lone parenthood and disability fields were very low, suggesting that the Medical Card Register is a reliable dataset. Modifications to the chronic disease score for the Irish setting were described and tests of the effect of GP prescribing style on measurement of chronic disease scores were outlined. We found that its effect was very small for all chronic illnesses, with the possible exception of cardio-vascular disease. GP variables, including a new set of indicators of prescribing style, were then assessed. We excluded GPs that had a significant change in panel size over part of the study period. We also found that the GP of choice indicator was dubious and chose not to use it. Instead the prescribing GP indicator was used. Consequently, we do not know if the prescribing style indicators relate to the prescribing GP or the GP of choice. This is the same GP in most but not all cases. Thus, a small but manageable ambiguity in the interpretation of the indicators of prescribing style is created. Finally this section examined descriptive statistics.

Section 3.3 assessed the external validity by comparing the mean values for a number of socio-economic variables from the Medical Card Register with those for similar variables from CSO publications and found them to be reasonably similar. It also compared age-related prevalence of chronic illnesses used in this study with those from UK epidemiological studies and found that the former were lower than the latter, but that the gap closed with increasing age. However, it is unlikely that chronic disease scores could be used as a substitute for epidemiological estimates of community morbidity, whatever about their role as a complement to them.

Section 3.4 described the treatment of missing values in the database. While the default option for most statistical packages is listwise deletion, we applied multiple imputation, which produced a more efficient model and changed coefficients, in some cases by a considerable amount.

In section 3.5 collinearity of covariates was assessed and found to be low. Section 3.6 included the prediction of income for all individuals in the dataset, to assist in model evaluation in later chapters. We also identified a number of vulnerable groups, which will also assist in model evaluation. Both the predicted income exercise and the identification of vulnerable groups may also be of interest to policy-makers and future researchers in the area.

4. EMPIRICAL EXAMINATION OF OUTSTANDING ISSUES IN MODEL SPECIFICATION

This chapter has two objectives. First, the appropriate specification of the relationship between age and health utilisation is unclear from the empirical literature. We examine a number of specifications in order to choose the most suitable one. Second, as outlined in Chapter 2, individual-level health utilisation data are characterised by non-negativity, a high proportion of zeroes and a right-skew. However, Chapter 2 found that an additive model of health care utilisation, estimated using OLS, was easily the most popular method for risk-adjustment. We explore the applicability of the additive linear model to prescribing expenditure in Ireland given these features of the dataset.

There are three strands to the second objective. First, we use outlier identification techniques to assess the extent of outliers in the dataset, and outlier-robust regression to assess their effect on OLS estimates. Second, we use quantile regression techniques to examine the effect of the set of covariates on prescribing expenditure across the conditional distribution, rather than relying on their effect at the mean only, which is especially relevant if the mean is a poor measure of central tendency, as is often the case with skewed datasets. Quantile regression also serves a complement to the outlier analysis. Third, recent studies (Deb and Trivedi, 1997 and 2002; Deb and Holmes, 2000; Deb and Burgess, 2002; Jimenez-Martin et al., 2002) highlight the usefulness of finite mixture models to estimate the factors affecting health care utilisation, especially using count data. We examine whether or not there are a number of distinct sub-groups within the distribution of prescribing expenditure. The detection of such sub-groups would suggest the application of finite mixture models, thereby adding to the seven models that we identified in Chapter 2 for estimation. In all cases, the Supply model is our preferred empirical specification as it contains variables from all five categories of variable described in Chapter 2 and is therefore less likely to suffer from omitted variable bias.

Estimating finite mixture models are exceptionally time consuming. Therefore, we take a random sub-sample for the exploratory analysis. Deb and Burgess (2002) found that the number of observations required to stabilise prediction errors to their

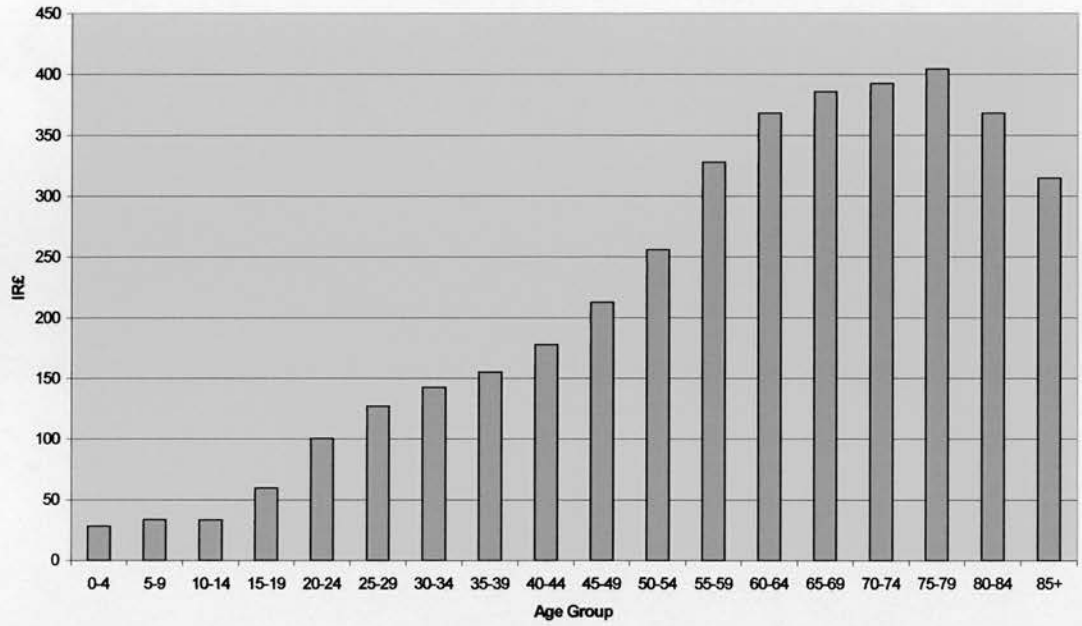
asymptotic values varied by estimator. OLS stabilised between 15,000 and 20,000 observations, while finite mixture models stabilised between 30,000 and 40,000 observations. Quantile regression and iteratively re-weighted least squares are also very time consuming. Consequently, we randomly chose a sample of 38,073 observations for exploratory analysis. No statistical differences between the sub-sample and the full sample were detected for any variable.

Section 4.1 examines the specification of age. Section 4.2 describes the distribution of prescribing expenditure and number of items prescribed. The next three sections describe aspects of this distribution. Section 4.3 describes and performs outlier analysis; section 4.4 describes and performs quantile regression and section 4.5 describes and performs finite mixture modelling. The dominant theme emerging from these three sections is that high cost patients differ from low cost patients. Thus approaches to the treatment of high cost patients are considered in section 4.6. Section 4.7 provides conclusions.

4.1 SPECIFICATION OF AGE

As with most health care utilisation data, the relationship between age and prescribing expenditure is non-linear. Figure 4.1 exhibits the average prescribing expenditure in five-year age groups (with one group for the over 85s, who form a group of about equal size to the others).

Fig 4.1 Bivariate Relationship between Prescribing Expenditure and Age



The relationship between age and prescribing expenditure is highly non-linear. It is characterised by low utilisation up to the age of 15, followed by a steady increase up to the forties, larger increases from there to the late sixties, a plateau in the seventies and a decrease for the over 80s. This pattern is similar to other types of health care utilisation, with the exception of the decrease in utilisation for those over 80. It is rare that clinical trials are done on the over 80s, so many expensive drugs are not prescribed to this age cohort. Those who survive to their eighties often do not have types of illness associated with high cost pharmacotherapy, such as coronary heart disease. This pattern of decreasing expenditure for people over 80 has also been observed in the UK (Lloyd et al., 1995). Indeed Joyce-Cooney (1999) called for this type of analysis to be done in Ireland, to see if the UK pattern existed in Ireland.

The current drug budget setting formula, based largely on the NARA, has age categories that are so wide that equivalence in prescribing expenditure is assumed between those aged 16 and 44, those aged 45 and 64, those aged 70 and 90. There are a number of parametric and non-parametric methods for specifying the relationship between age and prescribing expenditure, with different strands of the economics literature appearing to prefer different specifications.

Many risk-adjustment studies apply dummies for each age group. A recent Swedish study used only three such age groups (Andersson et al., 2000); the Dutch risk-adjustment studies routinely use nine (van Barneveld et al., 1997; van Barneveld et al., 1998; Lamers and van Vliet, 2001), while many studies choose up to 20 five-year age groups (Hornbrook and Goodman, 1995; Blough et al., 1999; Breyer, 2001). Not all risk-adjustment studies use age dummies, however. Holly et al. (2001) uses a quadratic in age.

Studies of the demand for health and health care are routinely use a quadratic in age specification (Pohlmeier and Ulrich, 1995; Cameron et al., 1988). Grossman (1972: 42) states *"since the curves relating [health and healthcare] to age would be concave to the origin... the square of age might be included as an additional explanatory variable. This variable should have negative coefficients in the demand curves for health and medical care"*. Moreover, a quadratic specification in age or experience has been routinely applied in human capital earnings functions since Mincer (1974). However, Murphy and Welch (1990) are highly critical of the specification, finding that it understated early career earnings growth by 30% - 50% and overstated mid-career earnings growth by 20% - 30%. Among their suggested alternatives are higher order polynomials.

A final alternative is the use of splines, which sets up the relationship between age and utilisation as a piece-wise linear function. The relationship between age and utilisation is graphed, the points at which there are obvious changes in the slope are identified by eye, and variables representing these slopes are generated. Using Figure 4.1, graphs using one-year age groups, as well as the average change from one age group to another, we decided to apply five splines with 'knots' (points where the slope changes) at 15, 45, 70 and 80. The knots can be expressed as dummy variables, D_k, where:

$D_k = 1$ if $age_{ij} \geq t_k$ or 0 otherwise, age_{ij} is the age of individual i attending GP j and $t_1 = 15$, $t_2 = 45$, $t_3 = 70$ and $t_4 = 80$. We can express the spline specification as:

$$y_{ij} = \beta x_{ij} + \phi z_j + \gamma age_{ij} + \sum_{k=1}^{k-1} D_k \lambda_k v_{ij} + \varepsilon_{ij}$$

where y_{ij} is prescribing expenditure of i attending GP j, x_{ij} and z_j is a row vector of other covariates included in the utilisation function, ε_{ij} is a independent, identically distributed error term, β, ϕ, γ and λ_k are coefficient vectors and $v_{ij} = age_{ij} - t_k$.

Therefore for those aged less than 15, age is represented as age_{ij} only. For those aged 15 to 44, age is represented by age_{ij} and their age less 15. For those aged 45 to 70, age is represented by age_{ij}, their age less 15 and their age less 45. The process continues for remaining ages.

We use OLS on the Supply model with cluster robust standard errors to relax the assumption of independence between observations attending the same GP. There are four different age specifications: five year age-dummies (with one for over 85s) as per Figure 4.1, denoted *'dum'*; a quadratic in age specification, denoted *'q'*; a cubic in age specification, denoted *'cub'*, and the five splines outlined above, denoted *'spl'*. Since the goal of risk-adjustment is ultimately to produce unbiased and efficient predictions, the criteria we apply to the selection of the specification of age are out-of-sample $\overline{R}^2$, root mean squared prediction error (RMSE) and Akaike Information Criterion (AIC). The exploratory dataset of 38,073 is randomly split into an estimation sample and a prediction sample and the three performance statistics is calculated for each specification. The process is repeated 100 times in what amounts to bootstrapping without replacement. The results are presented in Table 4.1.

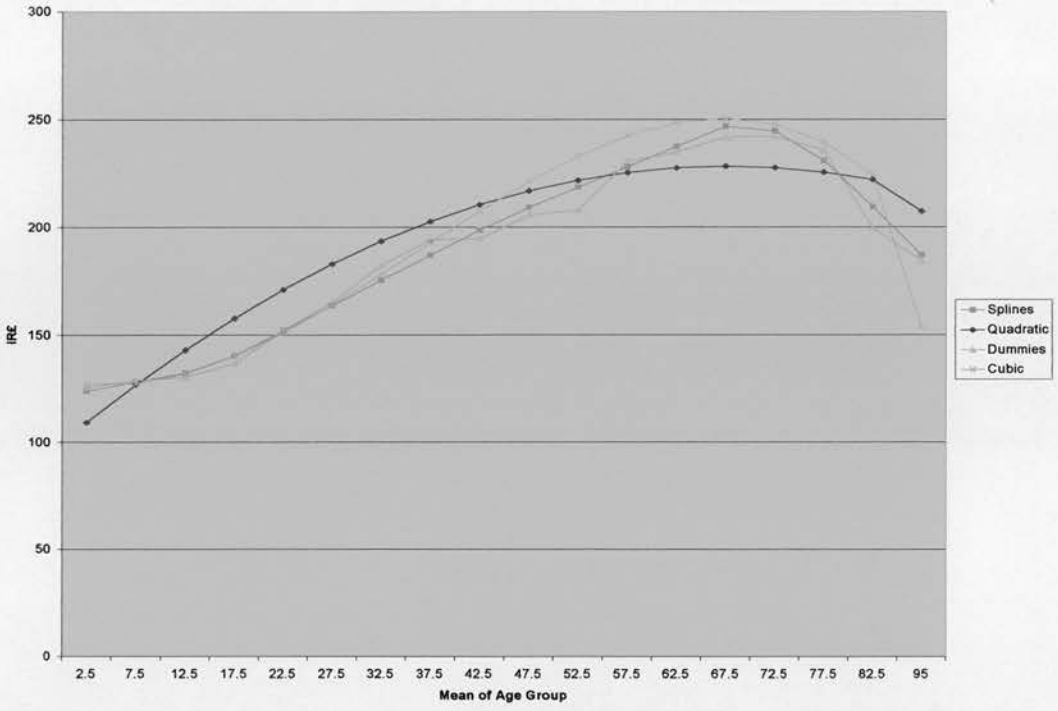
Table 4.1

Performance Statistics by Specification of Age

Specification	n	Mean	Std. Dev.	Min.	Max.
$\overline{R}^2$ dum[1]	100	29.12	1.12	26.40	32.08
$\overline{R}^2$ q[1]	100	29.11	1.11	26.37	32.07
$\overline{R}^2$ cub[1]	100	29.19	1.12	26.47	32.18
$\overline{R}^2$ spl[1]	100	29.19	1.12	26.47	32.15
RMSEdum	100	382.60	10.80	355.85	407.59
RMSEq	100	382.64	10.78	356.04	407.56
RMSEcub	100	382.42	10.81	355.71	407.36
RMSEspl	100	382.41	10.81	355.67	407.35
AICdum	100	11.87	0.06	11.73	12.01
AICq	100	11.87	0.06	11.73	12.01
AICcub	100	11.87	0.06	11.73	12.01
AICspl	100	11.87	0.06	11.73	12.01

1. Reported as a percentage

Figure 4.2 Comparison of Age Specifications by Age Group



The overriding result from Table 4.1 is the equivalence of specification of age for each performance statistic. The average $\overline{R}^2$ varies from 29.11% for the quadratic specification to 29.19% for the cubic and spline specification. The variance statistics show that we cannot distinguish between the four specifications. Similarly, the lowest root mean squared prediction error is the spline specification with the highest being the quadratic specification, but we cannot distinguish between them statistically. Finally, all specifications produce the same results for the Akaike Information Criterion, except at very high numbers of decimal places. In conclusion, we cannot distinguish statistically between the four age specifications.

The equivalence of age specification is further illustrated in Figure 4.2, which graphs the predicted value for each specification by age (controlling for other covariates except the intercept). Midpoints of the age groups in Figure 4.1 are used.

The spline and dummy specifications are practically indistinguishable in all but a few age groups. This is as expected since both are based on the empirical data. The cubic specification follows the spline and dummy specification up to the late twenties, then increases more rapidly than them up to age 57.5, converging upon

them again in the mid-60s, followed by a sharper decrease for those aged over 82.5. The quadratic specification produces the smoothest locus, with higher expenditures for those aged between 7.5 and 42.5 than the other three specifications, and lower expenditures for those aged between 57.5 and 77.5 than the other 3 specifications. However, the overriding result from Figure 4.2 is that the four age specifications are equivalent. We could include confidence intervals for each specification, but what we would gain in statistical clarity we would more than lose graphically.

In conclusion, a number of different specifications of age have been applied in empirical work on health care utilisation. Meanwhile, Murphy and Welch (1990) are highly critical of the quadratic specification, which is the most popular in empirical applications of human capital earnings function. Unlike Murphy and Welch (1990) – but for slightly different performance criteria – we find that all specifications of age are statistically equivalent. The quadratic approach is by far the most established in empirical work on human capital theory, including applications of the Grossman model. In addition, it is simple to interpret and does not rely on arbitrary knots like the spline model and is our choice in subsequent models.

4.2 DISTRIBUTION OF PRESCRIBING EXPENDITURE

As described in Chapter 2, health utilisation data is characterised by a large proportion of zeroes and a right skew, with some individuals generating much higher expenditure than the sample's 'typical' individual.

Table 4.2 describes the distribution of prescribing expenditure for the full sample of 400,751 observations, the number of items prescribed and the average cost per item at each of the selected expenditure percentiles. In addition, one of the mechanisms used by the IDTS to prevent GPs from being overexposed to high cost patients is by declaring certain drugs as 'budget neutral' as described in Chapter 3. Table 4.2 also describes expenditure on 'budget neutral' drugs and expenditure with 'budget neutral' drugs removed.

Table 4.2

Distribution of Prescribing Expenditure with and without Budget Neutral Drugs

Pctile	All Prescribing Expenditure	No. of Items Prescribed	Avg Cost per Item Prescribed	Expenditure of Budget Neutral Drugs	Expenditure with Budget Neutral Drugs removed
Min	0.00	0	0.00	0.00	0.00
1st	0.00	0	0.00	0.00	0.00
5th	0.00	0	0.00	0.00	0.00
10th	0.00	0	0.00	0.00	0.00
25th	2.01	1	2.01	0.00	2.01
50th	30.89	6	5.15	0.00	30.89
75th	216.37	29	7.46	0.00	216.37
90th	643.23	32	20.10	0.00	643.23
95th	1,016.38	67	15.17	25.17	991.21
99th	2,066.48	176	11.74	318.38	1,748.10
Max.	17,202.33	687	25.04	495.04	16,707.29
Mean	211.42	18.97	11.15	15.26	196.16

The degree of skewness is vividly demonstrated in the second column. Twenty three percent of the total sample are non-users in the study period, while the expenditure on the median individual is only IR£30.89 compared with a mean expenditure of IR£211.42. While expenditure on 95% of the sample is less than IR£1,017, expenditure rises sharply thereafter. Between the 95th and 99th percentiles, expenditure rises from IR£1,016 to IR£2,066, while expenditure on the top 1% varies from IR£2,066 to IR£17,202.

The number of items prescribed increases with percentile of expenditure. For instance, at the 99th percentile of expenditure 176 items are prescribed. However, the number of items prescribed does not increase as fast as total expenditure, meaning that the average cost per item prescribed increases with increasing expenditure. The average cost per item above the 95th percentile of expenditure is IR£18.54 as against IR£8.77 at or below the 95th percentile. Above the 99th percentile, average cost per item is IR£26.78 as against IR£10.16 at or below that percentile. Therefore, the high expenditures for some patients are a combination of more items prescribed and a higher cost per item.

The fifth column shows the cost of budget neutral drugs by percentile of overall expenditure. We can see that although only high cost patients are indicated as getting budget neutral prescriptions, these prescriptions do not amount to a large fraction of their total expenditure. For instance, that patient at the 95th percentile had total expenditures of IR£1,016.38, but the expenditure on their budget neutral drugs

was only IR£25.17. Therefore, we can see in the sixth column that there are still many high cost patients even when budget neutral drugs are removed. Clearly the budget neutral scheme is quite modest and does not alter skewness of the distribution of prescribing expenditure to a great degree. Indeed, the third moment for total expenditure is 5.71, while it is 5.61 when budget neutral drugs are removed, indicating that skewness is not altered dramatically by the removal of budget neutral drugs.

4.3 OUTLIER IDENTIFICATION

Assuming a linear additive model, it is likely that the high cost patients described above can be considered 'outliers'. We must exercise caution, however, in considering an observation to be an outlier as we may be throwing away valuable information. After all, it is possible to 'prove' that planets have a circular orbit of the sun by removing those planets at the further points on their elliptical orbit. Ideally we would like to be able to show that certain observations are from a different data generating process than the majority of observations, that is, their utilisation of health care is structurally different from other observations. If, for instance, certain observations were participants in a clinical trial, or if they could not speak English, then we could safely say that they are outliers, and can be treated differently from the majority of observations. While it is unlikely that we can prove that certain observations are from a different data generating process, we can highlight those observations that have an undue influence on coefficient values, and which should be considered for removal from that model applied to the majority of observations.

There is a growing literature on the principles and application of outlier identification and outlier-robust estimation. Meanwhile treatment of outliers in the risk-adjustment literature does not appear to have adopted these principles. Hornbrook et al. (2001) is the only recent example of a risk-adjustment study outlining its treatment of outliers. They truncated their data at various levels until the influence of outliers was considered acceptable, which in their instance was to remove 15 patients with annual expenditures of greater than IR£400,000 out of a sample of 1.5 million. Their assessment of what an acceptable level of influence was unstated.

Although the rigorous identification of outliers has not been published in the health care utilisation literature, there are examples of such studies elsewhere is empirical economics. Zaman et al. (2001) compared the results of OLS and robust estimates for two growth models and a stock return model, where least trimmed squares was used for robust regression. They found that in all cases, the explained variance improved and the value of coefficients and significance of variables changed when the robust model was applied.

4.3.1 Methods

In order to identify outliers, one needs to consider both residuals and leverage, where leverage is the distance from an observation's x_i to the centre of all the x_i's. For instance if the majority of observations are concentrated in the south west corner of the xy space with one in the north east corner, then that observation in the north east may have a small residual, in that the regression line may pass directly through it, but it has high leverage. Leverage points exercise a considerable effect on slope coefficients. We treat as outliers observations with high leverage and a large residual, so-called bad leverage points.

We employ a type of generalised M-estimator for robust regression (Goodall, 1983). While OLS has a 0% breakdown value, meaning that a small percentage of deviant observations can change coefficients to any value from -∞ to +∞, generalised M-estimators can have breakdown values of up to 30%, decreasing with the number of parameters in the model (Rousseau and Leroy, 1987). We detect outliers using Cook's D (Cook, 1977), which considers the observation's residual and its leverage, expressed as

$$D_i = \frac{1}{k}\frac{s_{(i)}^2}{s^2} r_i \sqrt{\frac{h_i}{1-h_i}} \qquad (4.1)$$

where D_i is Cook's D, k is number of parameters, s is the root mean square error of the regression, $s_{(i)}$ is root mean square error when the i^{th} observation is omitted, r_i is the studentized residual of the i^{th} observation and h_i is the leverage of the i^{th} observation.

We employ iteratively re-weighted least squares to generate robust regression estimates. First we run OLS, calculating Cook's D and dropping observations for

which D > 1. We then calculate case weights based on absolute residuals, regress again using those weights and continue the process until convergence, where a change in case weights of less than 0.01 is the convergence criterion. Weights are calculated using two weight functions, namely Huber weights (Huber, 1964) and biweights (Beaton and Tukey, 1974). First, Huber weights are used until convergence and, based on that result, biweights are used until convergence. Both weighting functions are used because Huber weights have difficulty dealing with severe outliers, while biweights may fail to converge or produce multiple solutions. Therefore, the initial Huber weights allow the behaviour of the biweighting to improve.

To derive Huber weights, first a scaled residual is calculated as

$$u_i = \frac{e_i}{s} \tag{4.2}$$

where u_i is the scaled residual, e_i is the i^{th} residual and $s = \frac{M}{0.6745}$, the residual scale estimate. M is the median absolute deviation from the median residual, or $M = med(|e_i - med(e_i)|)$.

Huber weights, (w_i) are set to 1 if $|u_i| \leq c_h$. Otherwise they are set as $c_h/|u_i|$. We set c_h to 1.345, which means that downweighting occurs if

$$|e_i| > \frac{1.345M}{0.6745} \tag{4.3}$$

$> 2M$ approximately, that is, downweighting occurs when the absolute residual exceeds twice the median absolute deviation from the median residual.

The biweights function sets w_i to 1 if residuals equal zero; all other residuals are assigned a weight of $\{1-(u_i/c_b)^2\}^2$ if $|u_i| \leq c_b$ and a weight of zero otherwise. We set $c_b = 4.685$. Therefore, a weight of zero is applied if

$$|e_i| > \frac{4.685M}{0.6745} \tag{4.4}$$

$> 12M$ approximately.

While we choose 12M as a threshold, anywhere between 6M and 12M is considered reasonable. The closer to 12M one chooses, the less the degree of downweighting

and the more like OLS the results become. Therefore, we are being conservative in our definition of an outlier. Standard errors are calculated as pseudovalues (StataCorp, 2000).

4.3.2 Results

We identify 2,044 observations (5.3%) as having a Cook's D of greater than 0.000105 (4/n), which is the usual cut-off (StataCorp, 2000). The distribution of prescribing expenditure for these 'outliers', as well as 'non-outliers' is described in Table 4.3.

Table 4.3

Distribution of Prescribing Expenditure for 'outliers' and 'non-outliers'

Statistic	'Outliers' IR£	'Non-Outliers' IR£
Minimum	0.00	0.00
5%	38.86	0.00
25%	345.14	1.80
50%	1,331.55	28.18
75%	1,865.62	184.92
95%	3,108.51	745.83
Maximum	12,542.78	1,837.57
Mean	1,349.07	152.93
N	2,044	36,029

Those identified as 'outliers' are high cost patients in general, with an mean expenditure of IR£1,349.07 as against IR£152.93 for 'non-outliers'. The difference in expenditure is even greater at the median. However, from the minimum and 5th percentile values we can also see that low cost individuals can also be identified as 'outliers'. These are individuals who may have high predicted expenditure and low actual expenditure, leading to a large residual, or whose characteristics in the set of covariates differ from the norm, meaning that they are high leverage points.

Table 4.4 describes the distribution of weights applied to observations for the robust regression estimates to follow.

Table 4.4

Case Weights for Robust Regression

Statistic	Weight
5th percentile	0
10th percentile	0.03
25th percentile	0.84
50th percentile	0.98
75th percentile	1.00
90th percentile	1.00
Mean	0.82
Number of observations	38,073

In total 3,423 observations (9.0%) get a weight of zero and are effectively dropped from the analysis. The average expenditure of patients with zero weight is IR£1296.12, as against an average expenditure for the rest of the group of IR£110.56, illustrating that those observations identified as 'outliers' are usually high cost patients. Indeed, many more high cost observations are identified as 'outliers' using robust regression than using simply Cooks D. Using robust regression, the first percentile of expenditure is IR£419.98 so there are a tiny percentage of low expenditure individuals identified as 'outliers'. Thus, robust regression refers to an average annual prescribing expenditure of IR£110.56, as against an average of IR£217.14 for OLS regression.

Further differences between observations assigned a weight of zero and those with a higher weight are described in Table 4.5, which compares 'outliers' and 'non-outliers' for each variable, ranked by 'outlier' value as a multiple of 'non-outlier' value.

Table 4.5

Comparison of 'Outliers' and 'Non-Outliers' for each Variable

Variable	'Outliers' (O)	'Non-Outliers' (NO)	O/NO
Totcost	1288.63	110.11	11.70
Rheum	0.10	0.02	6.44
Comor	0.42	0.07	5.78
Diabetes	0.07	0.01	4.85
Psych	0.21	0.05	4.16
Respir	0.17	0.04	4.11
Epi	0.05	0.01	3.84
Glau	0.03	0.01	3.83
CVD	0.47	0.14	3.24
Thyroid	0.05	0.02	2.69
Disabil	0.19	0.11	1.73
Agesq	4275.24	2541.73	1.68
Age	62.93	42.75	1.47
5-7 miles	0.21	0.20	1.05
3-5 miles	0.15	0.14	1.05
SHB	0.43	0.41	1.04
Specific	0.48	0.47	1.01
Gpagesq	2635.12	2615.60	1.01
Symptomatic	0.20	0.20	1.01
HB	5.91	5.87	1.01
Decpanel	932.06	928.28	1.00
Dist	1.72	1.71	1.00
Nurse	0.69	0.69	1.00
GPage	50.79	50.64	1.00
Sec	0.92	0.92	1.00
Rural	0.62	0.61	1.00
10+ miles	0.01	0.01	0.99
Presum	0.17	0.17	0.99
d1	0.59	0.60	0.98
SEHB	0.31	0.31	0.98
Gender	0.55	0.56	0.97
Marital	0.48	0.50	0.97
NEHB	0.26	0.27	0.96
RPA	0.12	0.14	0.87
7-10 miles	0.03	0.03	0.79
Numcard	1.62	2.54	0.64
Lonepare	0.02	0.11	0.23
N	3,458	34,615	-

The clear difference between the 'outliers' and 'non-outliers' is the level of prescribing expenditure, with 'outliers' costing almost twelve times as much as 'non-outliers', on average. The next set of notable results are for the chronic disease indicators, where observations indicated as 'outliers' are much more likely to be indicated as having chronic diseases than 'non-outliers'. Whereas only two percent of 'non-outliers' have rheumatological conditions, ten percent of 'outliers' do, a 6.4 fold difference. Forty two percent of 'outliers' have comorbidities as against seven percent of 'non-outliers', a 5.8 fold difference. As well as chronic disease indicators, disability and age exhibit considerable differences between the 'outliers' group and

the 'non-outliers' group. While 11% of 'non-outliers' are indicated as disabled, 19% of the 'outliers' are indicated as such; while the average age of 'non-outliers' is 42, it is 62 for 'outliers'. On the other hand, 'outliers' are less likely to be lone parents and are less likely to come from a large family, as indicated by the number of household members on the medical card. These results indicate considerable heterogeneity in prescribing expenditure for those indicated as having chronic diseases, those indicated as coming from households in receipt of disability payments, while heterogeneity increases with increasing age. Heterogeneity in the relationship between prescribing expenditure and each of these variables is explored further using quantile regression below.

The OLS and outlier-robust specifications of the factors affecting prescribing expenditure, using the Supply model, are presented in Table 4.6. Heteroscedasticity robust standard errors are applied.

Table 4.6

Determinants of Prescribing Expenditure: OLS and Robust Regression Compared

Variable	OLS		Robust	
	Coeffic	P>|t|	Coeffic.	P>|t|
Age	3.79	0.00	1.59	0.00
Agesq	-0.03	0.00	-0.01	0.00
Gender	-6.78	0.09	6.81	0.00
Disabil	69.43	0.00	10.78	0.00
Lonepare	-2.21	0.64	5.84	0.00
Marital	18.50	0.00	5.49	0.00
Rural	-8.34	0.10	-0.97	0.40
Numcard	-13.23	0.00	-1.76	0.00
CVD	306.45	0.00	190.82	0.00
Epi	252.94	0.00	96.66	0.00
Rheum	344.14	0.00	120.55	0.00
Diabetes	367.50	0.00	214.28	0.00
Glau	293.49	0.00	249.34	0.00
Respir	357.05	0.00	140.56	0.00
Thyroid	116.82	0.00	45.65	0.00
Psych	371.93	0.00	160.32	0.00
Comor	-22.26	0.22	12.57	0.00
3-5 miles	8.73	0.19	-2.59	0.07
5-7 miles	1.02	0.86	-4.17	0.00
7-10 miles	-22.68	0.03	-2.86	0.27
10+ miles	4.47	0.79	-3.46	0.37
NEHB	27.02	0.00	2.34	0.04
SEHB	16.74	0.00	2.16	0.06
GPage	-10.09	0.00	-1.91	0.01
Gpagesq	0.10	0.00	0.02	0.01
Nurse	13.93	0.00	4.33	0.00
Sec	-6.71	0.36	-2.71	0.12
Decpanel	0.00	0.87	0.00	0.42
Specific	130.86	0.17	79.29	0.00
Symptomatic	338.79	0.00	176.79	0.00
Presum	239.55	0.09	166.35	0.00
RPA	-4.67	0.48	-6.58	0.00
_cons	100.30	0.36	-50.23	0.05
Adjusted R^2 (%)		29.49		61.2
N		38,073		38,073

The reference categories for the distance dummies and the Health Board dummies are less than 3 miles from the GP and the Southern Health Board respectively, as these best approximated the mean prescribing expenditure for the entire population.

The adjusted R^2 for the OLS model is 29.42%, which is high for most cross-sectional studies and especially studies of the demand for health care. Prescribing expenditure is found to be quadratic in age. The disability variable is strongly positive, as is marital status. The distance dummies are jointly insignificant, as is living in a rural area, gender and lone parenthood. Each of the health board dummies are significant. All chronic disease indicators are strongly positive, as expected, with the exception

of the comorbidities indicator, which is insignificant. Prescribing expenditure is quadratic in GP's age and strongly negative for the range of GP's ages. Having a practice nurse and the proportion of total prescribing that is symptomatic are additional significant supply-side variables. A one percent increase in the percentage of prescribing that is symptomatic leads to a IR£3.39 increase in the patients prescribing expenditure. All other supply-side variables are insignificant.

The robust model produces a much better fit than the OLS model, with an R^2 of 61.2%. The coefficients in general are more consistent with expectations and are more strongly significant, suggesting a more efficient specification. Since the robust model is regressing about a lower mean, coefficients are lower than their corresponding OLS values in general.

Like the OLS model, it is quadratic in age. Gender is both significant and strongly positive, as expected. Disability has a much smaller effect in the robust model, as a higher percentage of those with a disability were assigned a weight of zero than of the remainder of the sample. The distance dummies are jointly significant and negative, as expected. The number of people on the medical card is again negative, but has a much smaller coefficient than in the OLS specification. The health board effects are less important in the robust model, with only the North Eastern Health Board being significant and with a much smaller coefficient. The coefficients on the chronic disease indicators are about half of their OLS values, although rheumatology is only 35% of its OLS value, and respiratory illness and thyroid disorders are less than 40% of their OLS values. Glaucoma, on the other hand, is 85% of its OLS value. As well as often presumptive and symptomatic, the same supply-side variables are significant with the same signs as in the OLS model, with smaller coefficients than in the OLS model.

4.3.3 Discussion

The most common model for risk-adjustment is an additive model, estimated using OLS. However, health care datasets are positively skewed with a long right tail, meaning that some observations can bear an undue influence on results. Moreover, OLS has a zero breakdown point, so it is particularly vulnerable to the biasing effect of outliers. While no risk-adjustment study or health care utilisation study that we

are aware of has used them, outlier identification techniques are now well established (Zaman, 2001). Rather risk-adjustment studies tend to either truncate the dataset at the 99.9th percentile (Deb and Burgess, 2002) or at some arbitrary level of expenditure (Hornbrook et al., 2001). We used both Cooks D and a generalised M-estimator to assess the extent of outliers in the dataset, as well as their effect on OLS estimates. We found that 'outliers' were generally high cost. The generalised M-estimator produced considerably different set of results compared with OLS.

Since we have no theory of why a particular observation is an 'outlier', we wish to be conservative in our treatment of observations. We do not want to drop observations that are not outliers. Hence, we postpone a discussion on what to do with outliers until section 4.6, after we have further explored the dataset using quantile regression and finite mixture models.

4.4 QUANTILE REGRESSION

Quantile regression is a useful way of describing the relationship between each covariate and prescribing expenditure across the conditional distribution of prescribing expenditure. We can use it to assess the effect of high cost patients on regression coefficients compared with low cost patients, in order to complement the outlier identification analysis in the last section.

Quantile regression is being used increasingly in labour economics to examine the returns to education (Hartog et al., 2001), or the effect of unionisation (Buchinsky, 1994), for instance. It has been applied previously in health research to resource use in rheumatology (Lambert et al., 1998), the covariates with birthweight (Abreveya, 2001), and the demand for alcohol (Manning et al., 1995).

4.4.1 Methods

Quantile regression is performed using least absolute deviation (LAD) regression, which minimises the sum of the absolute residuals rather than the squares of the residuals as in OLS (Koenker and Bassett, 1978; Koenker and Bassett, 1982; Koenker and Hallock, 2001). Taking q as the quantile being estimated, define a multiplier $h_i = 2q$ if $e_i > 0$; $2(1\text{-}q)$ otherwise. Using linear programming techniques,

we minimise the sum of weighted absolute deviations $\sum_{i=1}^{n}|e_i|h_i$ with respect to β.

Standard errors are estimated using the (x,y)-pair bootstrap, with 100 replications.

4.4.2 Results

The value of each of the covariates at the 20th, 40th, 60th, 80th and 95th quantiles, as well as the mean value based on the OLS model, is illustrated in the graphs in Figures 4.3, 4.4 and 4.5. Each graph represents one of the variables in the model. The dashed, straight line flanked by the two dotted straight lines are the OLS coefficients and confidence intervals, while the continuous, kinked line flanked by the two dashed, kinked lines are the quantile coefficients and confidence intervals. Figure 4.3 contains graphs of each of the demographic and socio-economic variables, while Figure 4.4 contains graphs of each of the chronic illness variables. Figure 4.5 contains graphs of supply-side variables and distance dummies. Graphs for the secretary variable (secretary), rural practice allowance (RPA) and panel size (decpanel) are not included, as they are insignificant in most models. Health board dummies and the intercept are also excluded. Since at the 20th percentile prescribing expenditure is zero, the coefficients for most variables for the quantile regression at the 20th percentile are zero.

Figure 4.3 Relationship between Prescribing Expenditure and Demographic or Socio-Economic Variables at Selected Quantiles

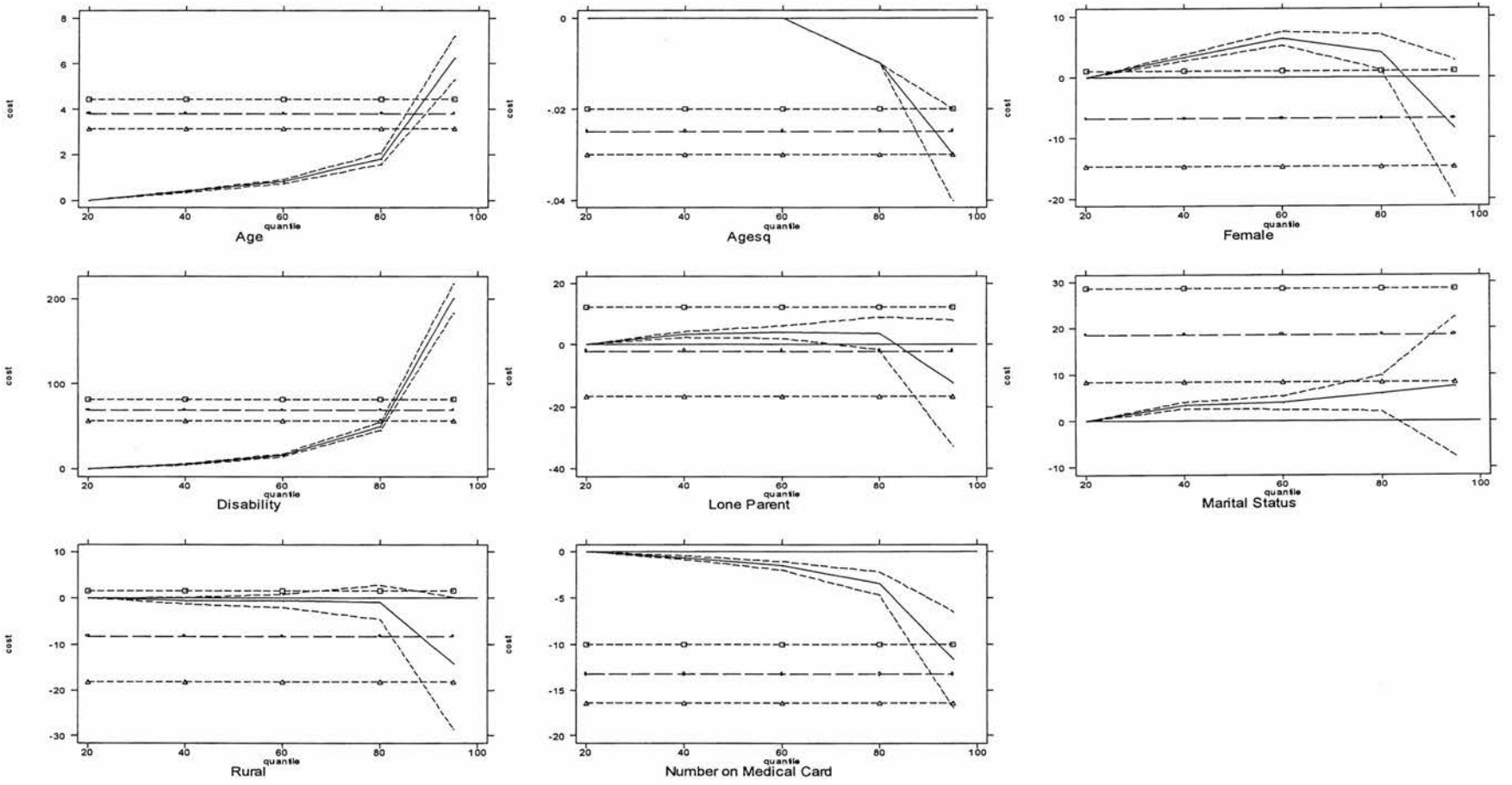


Figure 4.4 Relationship between Prescribing Expenditure and Chronic Illness at Selected Quantiles

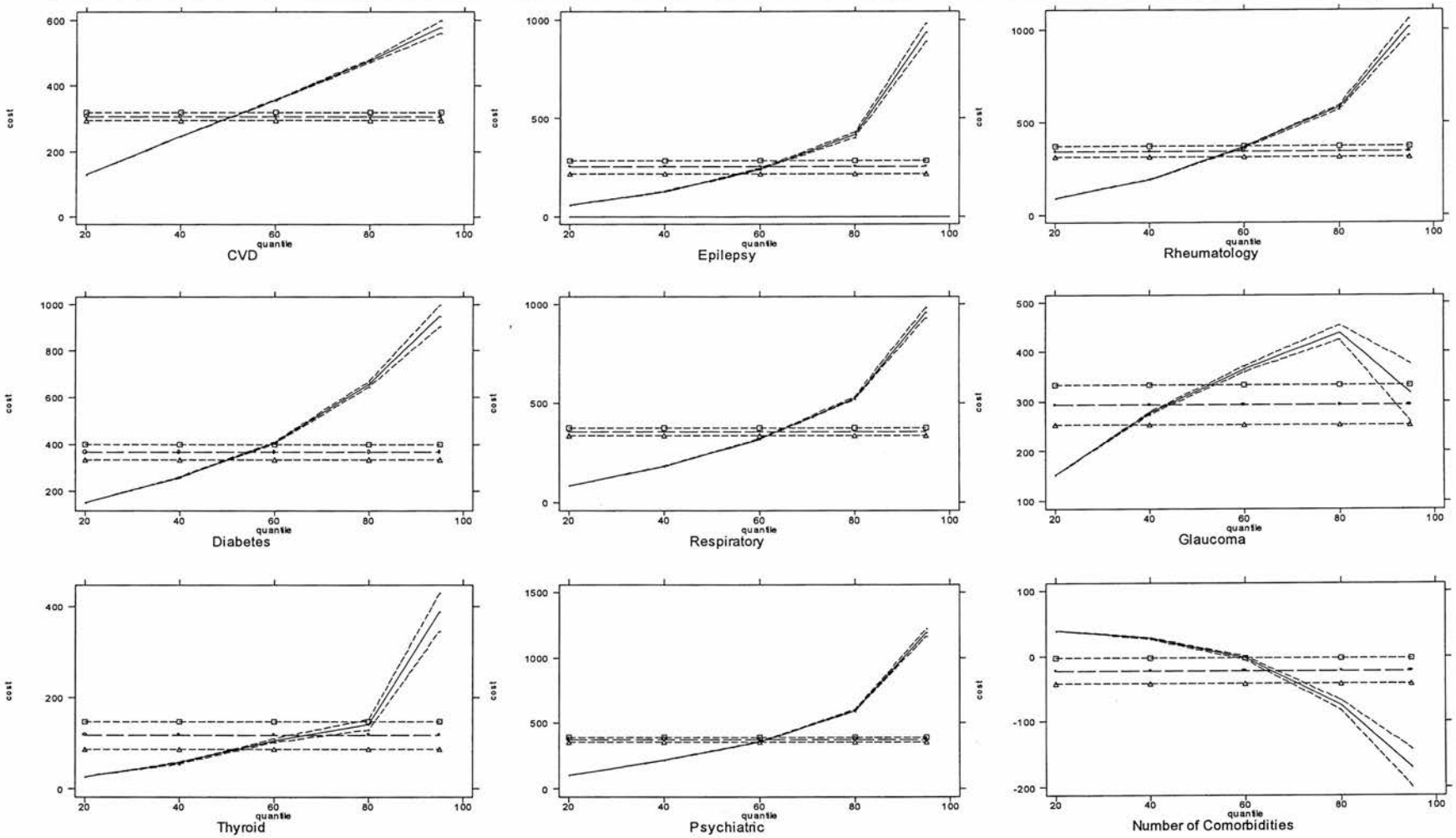
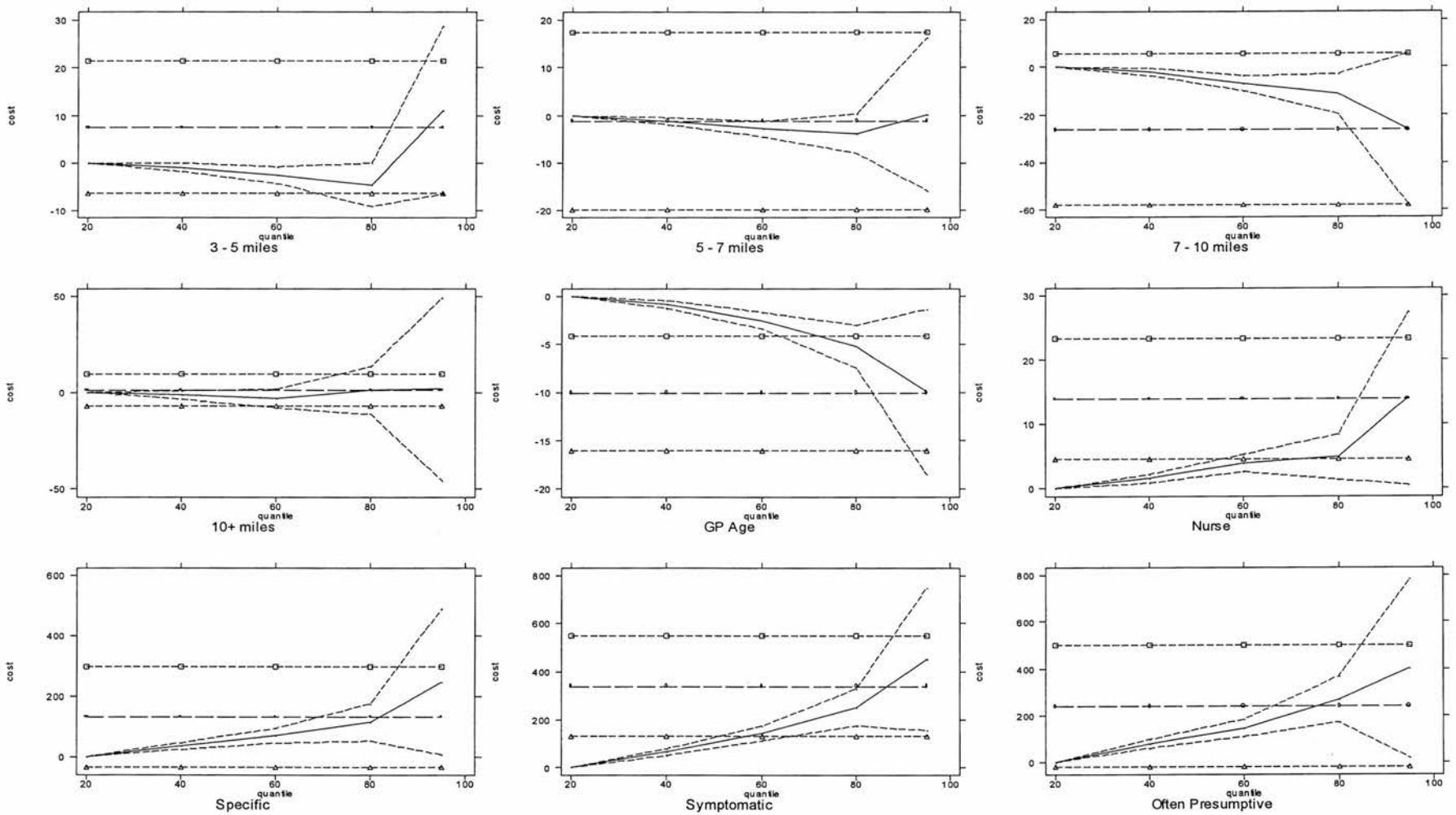


Figure 4.5 Relationship between Prescribing Expenditure and Distance or Selected Supply-side Variables at Selected Quantiles



The demographic and socio-economic indicators included in Figure 4.3 illustrate the degree of heterogeneity in the relationship between prescribing expenditure and each of the variables. A number of common patterns emerge. First, the relationship between prescribing expenditure and each of the covariates changes for each quantile; second, there is a sharp change in the relationship between the 80th quantile and the 95th quantile; third, the confidence intervals tend to fan out with increasing quantiles, with increasingly high cost users having a more variable relationship with the set of covariates, or in other words exhibiting heteroscedasticity related to expenditure. The heteroscedasticity does not seem to be directly related to any one covariate.

Age, age squared and disability all exhibit sharp positive kinks at the 80th to the 95th quantile. These findings accord with those in Table 4.5 above, where these three variables were found to be considerably affected by outlier identification and weighting. Sharp negative kinks are also found for females, lone parents, rural residence and number on the medical card, although only the last one of these is significantly different from zero at the 95th quantile. The OLS results imply that the expenditure associated with each additional year of age increases up to a maximum in the early 80s. For all but the 80th quantile, expenditure is increasing in age and age squared, while the 80th quantile reaches a maximum at 100 years of age.

Figure 4.4 exhibits the relationship between chronic disease indicators and prescribing expenditure by quantile. The coefficients on the chronic disease indicators vary even more by quantile than do the demographic and socio-economic variables, while the kink from the 80th to the 95th quantile is not as sharp (with the exception of epilepsy and thyroid disorders), and confidence intervals do not widen to the same degree. The OLS coefficients intersect the quantile coefficients at about the 60th quantile, as against the 80th quantile for most demographic and socio-economic variables. Unlike the other chronic illness indicators, the coefficient on glaucoma increases up to the 80th quantile and decreases thereafter, which accords with the finding in the robust regression above, where the robust coefficient for glaucoma was 85% of the OLS coefficient, as against 50% for all chronic disease indicators. The coefficient on number of comorbidities changes from positive up to the 60th quantile to negative thereafter.

Figure 4.5 exhibits the coefficients for each of the selected supply-side variables, as well as the distance dummies, by quantile. The robust regression analysis found that the number of observations that were a assigned a weight of zero did not differ much by any of the variables described in Figure 4.5. That finding is consistent with the patterns that emerge in Figure 4.5. Unlike in Figures 4.3 and 4.4, the OLS confidence intervals contain the quantile point estimates in the majority of cases. Despite this, the pattern of a kink between the 80th and 95th quantile and increasing variability by quantile remain.

Unlike in Figures 4.3 and 4.4, the coefficients on the distance dummies do not increase in value by quantile. Rather, they are insignificant in the majority of instances. GP age is increasingly negatively associated with expenditure by quantile, while the remaining variables included in the figure – nurse, specific, symptomatic and often presumptive prescribing proportions exhibit increasingly positive coefficients by quantile, along with increasing variability.

4.4.3 Discussion

This section applied quantile regression to the Supply model, in order to explore the relationship between prescribing expenditure and the set of covariates at different points on the conditional distribution of prescribing expenditure. We found that coefficients for many variables at selected quantiles were outside the OLS confidence intervals. For demographic and socio-economic variables, the OLS estimates intersected the quantile estimates between the 80th and 95 quantiles usually. For chronic illness variables, OLS estimates intersected quantile estimates at about the 60th quantile usually. Supply-side and distance variables were usually reasonably similar to OLS estimates, and usually within OLS confidence intervals. There is a sharp kink from the 80th to the 95th quantile for many variables, indicating again that for high cost patients the relationship between prescribing expenditure and the set of covariates is different to that for low cost patients and is often outside OLS confidence intervals. The case for their removal is discussed further in section 4.6.

A number of other interesting patterns emerge from this exploratory analysis that merit future research. For instance, it is unclear why older GPs have lower costs for

their high cost patients. Given the patterns in Figure 4.4, do high cost individuals who have chronic illnesses get higher cost chronic illness drugs than low cost individuals with chronic illnesses? Finally, why is disability such as strong determinant of prescribing expenditure for high cost individuals and less so for low cost individuals? Since the disability variable consists of long-term disabled (in receipt of invalidity pension), and short-term disabled, two groups are covered by this variable who could have very different prescribing needs. Although Schokkaert and van de Voorde (2000) found that people with a disability for more than one year had only 12% higher expenditures than people with a disability for less than one year, the separation of the disability variable may be a fruitful exercise. We leave it to future research to examine this and the other questions raised.

4.5 FINITE MIXTURE MODEL

As described above, outlier identification would ideally point out those observations for whom prescribing expenditure is generated through a different data generating process than the rest of the sample. The most common example of there being two 'types' within a health utilisation dataset is the two-part model, which splits the dataset into users and non-users and models each one separately, as described in Chapter 2. Deb and Trivedi (1997, 2002) point out that the two-part model implies that there is a mixture of distributions within the sample – one distribution for non-users and one for users. However, a more flexible approach would be to relax the assumption that separate distributions exist for users and non-users and examine how many distributions are mixed into the dataset and who belongs in which distribution. Finite mixture models can be used for this purpose.

Deb and Trivedi (1997, 2002), using finite mixture models found that rather than the sample splitting up into users and non-users, it split up into 'ill ' and 'healthy' sub-samples, where 'healthy' people were those who visited the physician occasionally or not at all and 'ill' people were those who visited frequently. They reason convincingly that the two sub-groups could be categorised as 'healthy' and 'ill', as it is health status rather than type of insurance cover or another covariate than distinguishes the two groups. Using data on health care expenditure from the RAND health insurance experiment, they found that these models performed better than

two-part models with respect to log-likelihood, AIC, Bayesian Information Criterion (BIC) and Andrew's Goodness of Fit (GoF) test.

As well as Deb and Trivedi (1997, 2002), finite mixture models have been used to estimate the demand for health care visits by Jimenez-Martin et al., (2002). Deb and Holmes (2000) and Deb and Burgess(2002) used the approach to analyse health care expenditures. In all cases the finite mixture model was compared to the two-part model, and in all cases the finite mixture model was favoured, with the exception of some of Jimenez-Martin et al.'s, (2002) findings, in particular that the finite mixture models outperformed the two-part model for GP visitation, but the two-part model was favoured for specialist visitation. They suggested that the multiple illness spell critique of the two-part model (Santos-Silva and Windmeijer, 1999) matters much more in the analysis of GP visitation than specialist visitation, which can be characterised adequately by the two-part model.

Deb and Holmes (2000) and Deb and Burgess (2002) are of most relevance to this study, as they model health care expenditures and concentrate on model prediction performance and implications for risk-adjustment. Deb and Holmes (2000) compared the finite mixture model for positive expenditures to the second part of the standard two-part model with a log-transformed response variable. The density of each component in the finite mixture model was assumed to be lognormal. Deb and Holmes (2000:487) state that *"it appears that the FMM constitutes a superior estimation strategy for setting capitation reimbursement rates because it can more accurately predict overall expenditures than standard estimation strategies"*. The study does not model zero expenditures and is based on a fairly small sample size (n=1594).

Deb and Burgess (2002) addresses the latter issue by using a sample of 2.5 million US patients. They compared eight competing models explaining health care expenditure: a linear OLS; a log-transformed OLS with a homoscedastic smearing estimate; a square-root transformed OLS; a linear OLS where negative predictions are set to zero; two generalised linear models based on the gamma distribution, one with a linear mean specification and one with a square mean specification, and finally two finite mixture models based on mixtures of gammas, one with two

components and one with three components. Their modelling strategy is similar in focus to that of Manning and Mullahy (2001), with four exceptions. First, Manning and Mullahy (2001) used logarithmic mean specifications for their generalised linear models, while Deb and Burgess (2002) found in preliminary work that the logarithmic mean specification performed poorly and chose to work with the linear and square-root mean specifications as a consequence. Second, Deb and Burgess (2002) include finite mixture models. Third, Manning and Mullahy include a heteroscedastic smearing estimate for retransforming logarithmic scale results. Fourth, Manning and Mullahy include additional generalised linear models with different variance functions, as described in section 2.2.3. Despite these differences, the broad strategy applied in both studies is similar.

Deb and Burgess (2002) find that the linear OLS model produces unbiased predictions but is relatively imprecise. They find that the GLM with a gamma distribution and a linear link function is relatively unbiased and produces better individual predictions, while the finite mixture model with a mixture of two gamma distributions performs better than the GLM model with linear link in most cases. If the ultimate goal is to minimise absolute prediction error, then the two component finite mixture model is probably the best approach to adopt, in their view.

4.5.1 Methods

Latent Gold 2.0 is used for finite mixture modelling, and the description of methods follows Vermunt and Magidson (2000). The response variable y_i is drawn from a mixture of sub-populations in proportions π_1, ..., π_C where $\sum_{j=1}^{C} \pi_j = 1, \pi_j \geq 0 (j = 1,...,C)$. The finite mixture model can be defined as:

$$f(y_i \mid \theta) = \sum_{j=1}^{C-1} \pi_j f_j(y_i \mid \theta_j) + \pi_C f_C(y_i \mid \theta_C) \qquad (4.5)$$

where θ is a set of parameters and $\pi_c = 1 - \sum_{j=1}^{C-1} \pi_j$. The product of the mixing probabilities π_j and the sub-population densities $f_j(y_i \mid \theta_j)$ is summed over the C sub-populations to give the overall density of y_i, while π_j and θ_i are estimated jointly.

Finite mixture models can be estimated by maximum likelihood or posterior mode methods. However, estimation is complicated by boundary solutions due to multinomial probabilities becoming zero or error variances converging to zero. We use a Bayesian approach. The former problem is overcome by using Dirichlet priors for the latent and conditional probabilities and the second problem is overcome by using inverse-Wishart priors for the error variance-covariance matrices. As a consequence of applying priors, the estimation method is Posterior Mode, rather than Maximum Likelihood. Essentially, Posterior Mode is a form of penalised maximum likelihood, in which the assumed priors for the set of parameters θ serves as a penalising function for solutions that are too near the boundary of the parameter space. Denoting the likelihood function log L, then the log-posterior function $\log P = \log L + \log h(\theta)$.

The Dirichlet prior for the latent probabilities equals $\log h(\pi_j) = \sum_{j=1}^{C} \frac{\alpha_1}{C} \log \pi_j$, where α is a user-defined Bayes Constant. The inverse-Wishart priors also apply Bayes constants. In both cases, we apply constants of one. The Dirichlet prior is equivalent to adding $1/C$ observations to each latent class, while the inverse-Wishart prior is equivalent to adding $1/C$ observations which are at a distance of one standard deviation from the class-specific mean and which have covariances of zero. With large sample sizes like that used in this study, the effect of priors on parameter estimates is negligible.

Since both π_j and θ_i are unknown, we use both EM and Newton Raphson algorithms. We begin with the EM algorithm, as described in section 3.4.2.1. Once the estimation procedure reaches the maximum number of iterations or converges we switch to the Newton Raphson algorithm, which then continues until convergence or until the maximum number of iterations is reached again. The maximum number of iterations is set at 1,000 for EM and 50 for Newton Raphson, while the convergence criterion is 0.1, that is, when the absolute sum of relative changes in parameter values in a single iteration is 10% or less. The stability of the EM algorithm when we are far away from the optimum motivates its choice as a starting algorithm, while the switch to the Newton Raphson algorithm is motivated by its speed.

In order to prevent local maxima, 10 sets of random starting values are applied. Twenty EM iterations are performed on each set of starting values, keeping the one with the best log-posterior and continuing until convergence. Since the EM algorithm is very stable, the use of multiple random starting values greatly increases the probability of finding a global optimum, although it cannot be guaranteed. Further details on finite mixture models are available in Vermunt and Magidson (2000) and McLachlan and Krishnan (1997).

4.5.2 Results

Table 4.7 exhibits model selection statistics for each of four finite mixture models.

Table 4.7

Finite Mixture Model Selection Statistics

	Log-likelihood	AIC
1 component	-282,838	565,743
2 component	-271,532	543,196
3 component	-264,482	529,162
4 component	-259,702	519,667

The 4-component model has the highest (least negative) log-likelihood and the lowest AIC. If we were to choose a model on purely statistical grounds then the 4-component model would be preferred.

However, the choice of model must also have a sound economic intuition. We examine the results of each model in the next table.

Table 4.8

Class Proportions and Mean Prescribing Expenditures for each Latent Class Model

Model		Class 1	Class 2	Class 3	Class 4
2-component	Class Size	0.97	0.03		
	Mean Prescribing Expenditure	178.34	2801.76		
3-component	Class Size	0.94	0.06	0.01	
	Mean Prescribing Expenditure	147.26	1252.60	4297.63	
4-component	Class Size	0.88	0.10	0.02	0.00
	Mean Prescribing Expenditure	125.31	581.49	1787.45	4881.43

The two component model places the majority of the sample (97%) into the first class. This class has an average prescribing expenditure of IR£178.34. The second

class is a high cost class, with average prescribing expenditure of IR£2,801.76. The first class of the three-component model splits high cost patients into two groups and retains 94% in the lowest cost class. As with other models, the four-component model places the majority of the sample into one low-cost group. Next, it places 10% of the population in a moderately high cost group, with an average cost of IR£581.49. It then places the remaining 2% of observations in two very high cost classes.

The results for a one-component model are simply the OLS results presented in Table 4.6 above. The results for the two-component model are described in Table 4.9. Although elsewhere in the text we report the p-value for statistical significance, we indicate those variables that are significant at the 5% level with an asterisk in this section, because the tables for three- and four-component models are large and would be too cluttered in p-values were included.

Table 4.9

Determinants of Prescribing Expenditure using a Two Component Finite Mixture Model

Variable	Class 1		Class 2	
	Coefficient		Coefficient	
Intercept	-13.18		11,585.05	*
Age	3.10	*	-115.25	*
Agesq	-0.02	*	0.80	*
Gender	6.20	*	316.08	*
Disabil	41.02	*	190.35	*
Lonepare	8.07		507.74	*
Rural	2.77		229.62	*
Numcard	-8.64	*	-83.55	*
Marital	24.33	*	265.87	*
CVD	257.72	*	23.64	
Epi	163.83	*	-255.79	*
Rheum	255.15	*	270.62	*
Diabetes	297.15	*	323.95	*
Glau	192.11	*	-80.69	
Respir	268.46	*	663.54	*
Thyroid	55.05	*	-174.43	*
Psych	264.54	*	536.11	*
3-5 miles	1.82		187.87	*
5-7 miles	-6.12		-29.68	
7-10 miles	-11.86		19.42	
10+ miles	-5.98		151.26	
NEHB	7.29	*	25.98	
SEHB	14.67	*	246.98	*
GPage	-8.03	*	-247.81	*
Gpagesq	0.08	*	2.29	*
Nurse	5.91		72.51	*
Sec	-2.44		-9.76	
RPA	-7.83		-113.35	*
Decpanel	0.00		-0.01	
Specific	186.10	*	498.22	
Symptomatic	344.64	*	1,310.43	*
Presum	332.10	*	-1,428.13	

The first class of the two-component model has a similar set of results to the OLS results outlined in Table 4.6 and discussed in more detail in Chapter 5. Indeed if OLS is truncated at the 97th percentile, the results are very similar indeed. Prescribing expenditure is quadratic in age, while being female is positively associated with expenditure. Disability, the number of people on the medical card and marital status are the three significant socio-economic variables, while lone parenthood and rural residence are insignificant. As with the OLS model in Table 4.6, there are large coefficients on the chronic illness indicators, ranging from IR£55.05 for Thyroid to IR£297.15 for Diabetes. Distance variables are insignificant, while on the supply-side only GP age (negatively associated with expenditure but at a decreasing rate over most GP ages) and prescribing style

variables are significant. The prescribing style variables indicate that, for instance, a 1% increase in the percentage of a GP's total prescribing expenditure that is often presumptive leads to an increase in prescribing expenditure of IR£3.32 for patients on that GP's panel.

The second class refers to 3% of individuals who have an average expenditure of approximately IR£2,800. The results for this class are more difficult to interpret. There is a large and significant intercept term. The large age coefficient suggests that age is has a strong negative association with prescribing expenditure. For instance, it implies that someone aged 45, which is approximately the sample average age, would have prescribing expenditure of IR£5,071 less than someone aged one, ceteris paribus. However, both the theoretical model in Chapter 2 and the bivariate analysis in Figure 4.1 suggest a positive relationship between age and expenditure for most ages at least. Moreover, the model outlined in Chapter 2 suggested a positive relationship between chronic illness and expenditure. However, there are negative coefficients on epilepsy and thyroid illness in this model. These results are particularly peculiar as Figure 4.2 found that at the 95th quantile there was a sharp increase in the coefficients on both variables. In addition, CVD and glaucoma are insignificant. There are also a number of coefficients with surprisingly high positive values, including gender, disability, marital status, rural residence and the South Eastern Health Board. The large positive coefficient on living 3-5 miles from the GP is also contrary to expectations. The results on the supply-side variables are also difficult to interpret. There are large coefficients on both GP age and GP age-squared. These results suggest that attending a 49-year old GP, which is approximately the sample average GP age, is associated with a reduction in expenditure of IR£4,419.36 compared with attending a 30-year old GP. There does not appear to be any economic intuition to the second class results.

Since there are results contrary to theoretical expectations on age and some chronic illness and distance indicators, and the coefficients are suspiciously large on other variables, it seems reasonable to conclude that the results for the second class are spurious. It appears that the two-component finite mixture model provides a consistent model for the lowest cost 97% of the sample but cannot model the top 3% of the sample.

Table 4.10 describes the results of a three-component finite mixture model.

Table 4.10

Determinants of Prescribing Expenditure using a Three Component Finite Mixture Model

Variable	Class 1		Class 2		Class 3	
	Coefficient		Coefficient		Coefficient	
Intercept	-185.35	*	411.78		-3,629.46	*
Age	2.62	*	-27.86	*	-120.23	*
Agesq	-0.02	*	0.18	*	0.87	*
Gender	6.69	*	56.47	*	505.97	*
Disabil	31.19	*	193.58	*	719.65	*
Lonepare	6.25		-121.94	*	5.87	
Rural	2.20		106.44	*	984.75	*
Numcard	-5.61	*	-75.04	*	-17.92	
Marital	18.53	*	106.09	*	284.76	*
CVD	233.01	*	257.41	*	-187.08	*
Epi	123.59	*	234.30	*	-50.85	
Rheum	182.73	*	241.62	*	164.72	*
Diabetes	241.08	*	280.83	*	18.66	
Glau	180.41	*	123.79	*	-783.59	*
Respir	208.37	*	435.93	*	754.99	*
Thyroid	56.57	*	-27.01		-1,236.51	*
Psych	212.92	*	442.76	*	733.31	*
3-5 miles	-6.04		-135.86	*	-611.39	*
5-7 miles	-4.49		11.49		1,125.79	*
7-10 miles	-21.01	*	-254.99	*	-963.24	*
10+ miles	-5.69		-107.02	*	-888.80	*
NEHB	7.38	*	120.50	*	141.98	*
SEHB	14.29	*	178.05	*	641.62	*
GPage	-3.01		-12.90		-78.71	*
Gpagesq	0.03		0.09		0.57	*
Nurse	4.47		62.09	*	565.38	*
Sec	-9.75	*	-213.59	*	-1473.77	*
RPA	-6.14		-1.66		-157.92	*
Decpanel	0.00		-0.07	*	-0.32	*
Specific	244.08	*	2,244.01	*	14,833.51	*
Symptomatic	378.93	*	2,750.37	*	12,548.60	*
Presum	428.77	*	2,681.62	*	22,139.74	*

The three-component model follows largely the same pattern as the two component model. The first class of the three-component model contains 94% of the sample with an average expenditure of IR£147.26. The results are very similar to OLS on the lowest cost 94% of observations. Coefficients are signed with expectations. The large and significant negative coefficient on the intercept is unusual by comparison with OLS, however.

In a pattern similar to the two-component model, the second class produces a number of counter-intuitive signs, chief of which is age. The age result implies that a 45-year

old would have prescribing expenditure of IR£1,225.84 less than someone aged one. In addition, there is a potentially counter-intuitive sign on lone-parenthood, but this is also counter-intuitive in OLS and is given a more detailed treatment in Chapter 5, where we find that a reasonable explanation exists. There are also some surprisingly large values on rural residence, some distance indicators and the presence of a GP's secretary. For instance, 91% of the sample attend a GP who has a secretary. The coefficient on this variable suggests that these individuals have prescribing expenditure that is IR£213.59 less than the 9% who attend a GP without a secretary. It seems reasonable to conclude that the results exhibit spurious associations.

The results for the third class, referring to just under 1% of the sample with an average prescribing expenditure of IR£4,297.63 also appear to exhibit spurious associations. Age is again negative, although age squared in positive. The combined effect is that a 45-year old would have IR£5,251.84 less prescribing expenditure than a one-year old, contrary to theoretical expectations. The negative signs on CVD, glaucoma, thyroid disease, as well as the positive sign on living 5-7 miles from the GP are also contrary to theoretical expectations. There are also a number of unusually large coefficients, such as GP age and GP age-squared. The combined effect of these two variables is that someone attending a 49-year old GP would have IR£1,484.66 less prescribing expenditure than someone attending a 30-year old GP. Other variables with unusually large coefficients include the rural residence, the last two distance indicators and attending a GP with a secretary, while the intercept is also unusually large.

Table 4.11 describes the results of a four-component finite mixture model.

Table 4.11

Determinants of Prescribing Expenditure using a Four Component Finite Mixture Model

Variable	Class 1		Class 2		Class 3		Class 4	
	Coefficient		Coefficient		Coefficient		Coefficient	
Intercept	-95.26		1,109.18	*	7,268.27	*	8,702.13	*
Age	2.10	*	18.43	*	-8.58	*	-165.99	*
Agesq	-0.01	*	-0.14	*	0.03		1.10	*
Gender	4.51	*	3.05		198.70	*	909.02	*
Disabil	22.09	*	86.45	*	160.73	*	-159.43	
Lonepare	8.07	*	-55.05	*	505.77	*	-6,314.71	*
Rural	-0.05		16.39		107.34	*	1,379.84	*
Numcard	-3.54	*	-65.20	*	-153.09	*	560.93	*
Marital	14.65	*	55.95	*	346.35	*	23.34	*
CVD	208.73	*	267.43	*	270.82	*	-399.49	*
Epi	105.15	*	277.87	*	195.32	*	-3,272.25	*
Rheum	146.84	*	274.70	*	230.36	*	-90.04	*
Diabetes	213.54	*	310.92	*	297.11	*	1,629.05	*
Glau	176.38	*	153.33	*	1.22		-5,879.44	*
Respir	189.00	*	455.11	*	777.19	*	472.14	*
Thyroid	65.37	*	135.98	*	573.81	*	-3,609.46	*
Psych	181.21	*	423.77	*	915.95	*	2,567.83	*
3-5 miles	0.25		47.66	*	295.04	*	-1,347.89	*
5-7 miles	-4.95		4.70		20.86		8,65.27	*
7-10 miles	-15.76	*	-64.96	*	-567.66	*	-1,920.42	
10+ miles	0.73		-16.65		-246.65	*	-1,505.19	
NEHB	5.51	*	57.98	*	193.74	*	-400.93	*
SEHB	7.54	*	38.39	*	127.71	*	1,015.54	*
GPage	-1.60		-0.58		-61.43	*	-0.10	*
Gpagesq	0.02		0.00		0.56	*	-0.11	
Nurse	2.74		4.44		67.23	*	956.72	*
Sec	1.13		41.70	*	73.46	*	-416.23	*
RPA	-6.07	*	-4.61		60.03		-836.56	
Decpanel	0.00		0.02		0.04		-0.14	
Specific	97.07	*	-1,248.23	*	-4,379.00	*	-1,239.59	*
Symptomatic	222.77	*	-642.21	*	-2,166.67	*	9,236.58	*
Presum	197.99	*	-1,857.45	*	-8,721.63	*	-14,629.51	*

The first class of the four-component model refers to the lowest cost 88% of observations, who have an average expenditure of IR£125.31. Age has a quadratic relationship with expenditure, increasing at a decreasing rate as expected. Other socio-economic and chronic illness variables exhibit similar coefficients as an OLS on the lowest cost 88% of observations. The distance indicators are insignificant except for living 7-10 more miles from the GP, which is signed as expected. Supply-side variables are insignificant, except for attending a GP in receipt of Rural Practice Allowance, which is negatively related to expenditure, and the prescribing style indicators, which are positively related to expenditure.

The second class of the four-component model refers to ten percent of observations with an average expenditure of 581.49. As such, they are moderately high cost

patients. The intercept is large, positive and significant. Age has a quadratic relationship with expenditure, where a 45-year old is expected to have expenditures of IR£804.76 greater than a one-year old. Disability and marital status are positive and significant, as expected, while rural residence is insignificant. Lone parenthood and number on the medical card are negative, which is consistent with OLS results, if unexpected *a priori*. Chronic illness indicators exhibit large positive coefficients as expected. Living 3-5 miles from the GP is positively associated with expenditure, which is contrary to theoretical expectations, while living 7-10 miles is negatively associated with expenditure as expected. Other distance indicators are insignificant. Health board effects are positive and significant. Supply-side variables are mostly insignificant, except attending a GP with a secretary, which is positively associated with expenditure and prescribing style, which are negatively associated with expenditure. In the case of symptomatic and often presumptive prescribing these are contrary to expectations, while a negative sign on specific prescribing is as expected. Overall, except for living 3-5 miles from the GP and the last two prescribing style indicators, the results are consistent with expectations.

The third class of the four-component model related to 2% of observations with an average expenditure of IR£1,785.45. There is a large positive intercept. Age has as strong negative, linear association with expenditure, contrary to expectations. There are large positive and significant coefficients on all other demographic, socio-economic and chronic illness indicators except glaucoma, which is insignificant. Unusually high coefficients obtain for the distance indicators. Living 3-5 miles from the GP is positively associated with expenditure, contrary to expectations, while living 7-10 miles and 10 or more miles form the GP have very high negative coefficients. It is difficult to believe that living 7-10 miles from the GP could cause a reduction in prescribing expenditure of IR£567.66. The GP age effect is quadratic, but again is very large. It implies that attending a 49-year old GP leads to a reduction in expenditure of IR£1,156.53 compared with attending a 30-year old GP. Prescribing style effects are negative, significant and very large. These negative coefficients are unexpected in the case of the last two. Of the remaining supply-side variables, attending a GP with a nurse and a secretary are positively associated with expenditure, while the size of the GP's panel and whether or not they received Rural

Practice Allowance are insignificant. We can safely conclude that the results are spurious.

The fourth class of the four-component model related to less than 1% of observations with an average expenditure of IR£4,881.43. The intercept is large, positive and significant. Expenditure is quadratic in age, with the large negative coefficient on the first polynomial dominating the overall effect. For instance, the combined age effects imply that a 45-year old would have prescribing expenditure of IR£7,255.16 less than a one-year old. The results are also unusual for a number of other demographic and socio-economic variables. The disability variable is insignificant; gender, rural residence and the number of people on the medical card have a surprisingly strong positive relationship to expenditure, while lone parenthood has a very high negative association. A number of the chronic illness indicators are also counter-intuitive, including cardio-vascular disease, epilepsy, rheumatology, glaucoma and thyroid illness. Living 5-7 miles from the GP is positively related to expenditure, contrary to expectations, while the coefficient on the 3-5 mile indicator is exceptionally high. The other distance indicators are insignificant. There is a modest, linear GP age effect. Attending a GP practice with a nurse has a large positive effect on prescribing expenditure, although attending a GP practice with a secretary has a large negative effect. The prescribing style effects are also very large. The negative coefficient on specific prescribing is as expected, as is the positive coefficient on symptomatic prescribing, while the negative coefficient on often presumptive prescribing is unexpected. The number on the GP's panel and rural practice allowance are again insignificant. Given the number of results that are contrary to expectation and the very high coefficients on a number of other variables, we conclude that these are spurious associations.

4.5.3 Discussion

The latent class models place the majority of individuals into one class– between 88% for the four-component model to 97% for the two-component model – which has below average prescribing expenditure. For all models, the coefficients in their first class are broadly consistent with theory and similar to OLS with high cost patients removed. The other classes, relating to high cost patients, produce counter-intuitive results and appear to be detecting spurious correlations. The exception is

the second class in the four-component model, which is consistent with theory for the most part. However, there does not appear to be any economic intuition for classes three and four of the four component model.

We can adopt one of two approaches based on these results. The results of the two-component model suggest that 97% of observations are in a group that have a stable and theoretically consistent relationship with the set of covariates, while the other 3% are high cost patients for whom the relationship between prescribing expenditure and the set of covariates is spurious. OLS truncated at the 97th percentile appears to be an acceptable modelling strategy based on the results of this model. The four-component model suggests that there are two groups that can be modelled – 88% of observations are in one low cost group and 10% of observations are in another moderate cost group. For both groups, the relationship between prescribing expenditure and the set of covariates is reasonably stable and consistent, moreso in the case of the lowest cost group. However, for the other two groups associations are spurious. In unreported work, we tried to model the pattern described by the four-component model as a two component finite mixture model truncated at the 98th percentile. This produced reasonable results for the first class and spurious results for the second class. The only way to produce a stable set of results for the two lowest cost classes is by using a four-component model.

The four-component model as an alternative to OLS is a little problematic, however. Finite mixture modelling is a useful exploratory technique, but it should only be applied when there is a sound reason for splitting the sample into distinct groups, that is, if some structural difference exists. If we have a set of observations, some of which belong to a control group in a clinical trial and others that belong to an experimental group, but we do not know which observations are which, a two-component finite mixture model could be used to separate observations into each group and they could legitimately be modelled separately. Deb and Trivedi (1997, 2002) and Jimenez-Martin et al. (2002) found that the two-component finite mixture model worked well for counts of health utilisation and they reasoned convincingly that such a dichotomy had sound economic reasoning, that is, it detected previously unobserved morbidity. Such a justification is absent from the results presented here, however.

Meanwhile, Deb and Burgess (2002) also found two and three component mixtures of gamma distributions had greater predictive ability than a number of competing models, including a model with a log transformed response variable and a GLM with a gamma density and a linear link function. They had no strong reasoning for why their sample should split into two or three distinct distributions, and OLS outperformed both these approaches for a number of criteria. They can potentially excuse the lack of justification for their mixtures by appealing to the size of their dataset, which at 2.5 million observations is likely to be very close to the actual distribution of population health expenditure for their group. A sound reason for the mixtures of distributions exhibited in these data would be a useful area of future research.

4.6 TREATMENT OF HIGH COST PATIENTS

Sections 4.3, 4.4 and 4.5 all identified differences between low cost patients and high cost patients in how their characteristics relate to prescribing expenditure. First, outlier identification and robust regression produced a broadly similar set of results and identified a significant number of 'outliers'. Cook's D identified 2,044 observations as 'outliers'. These had an average expenditure to IR£1,349.07 as against an average expenditure for non-outliers of IR£152.93. Moreover, robust regression assigned a weight of zero to 3,423 observations, meaning that 9% of the total sample was dropped from the regression estimates. Those observations with a weight of zero had an average expenditure of IR£1,296.12 as against IR£110.56 for other observations.

Second, quantile regression found a sharp kink between the 80^{th} and 95^{th} quantiles for all variables except marital status and cardiovascular disease, and to a lesser extent diabetes and the number of comorbidities. Third, finite mixture models classified between 88% and 97% of observations in one, low-cost, class. The two-component and three-component models, while having higher log likelihoods and lower AIC values than a one-component model, produced estimates for all classes except the first one that were not economically meaningful. Meanwhile, the four-component model did not produce meaningful estimates for its third and fourth components. Thus, finite mixture modelling suggested that high cost patients are

different to low cost patients and that they are not easy to model, at least in an additive linear framework. Therefore, all three exploratory techniques indicated that high cost patients were noticeably different from other individuals. Their coefficients had larger absolute values and they exhibited a greater degree of variability.

The key question is what to do with high cost individuals. We have not a developed theory of why some observations are outliers and others are not, or why there are two, three or four distributions in the dataset, rather than one. As such, we should be conservative in how we deal with 'outliers' or multiple distributions. However, to leave 'contaminating' observations in the model may have an undue influence on the results. The three options that we can consider are truncation, applying iteratively-reweighted least squares and applying a four-component finite mixture model.

The application of a four-component finite mixture model would produce spurious results for two of the four classes, and since we have no adequate theory for why the distribution should split into four classes, we do not wish to pursue this option. Finite mixture modelling has been useful here to point out again the difference between high cost and low cost patients but we are reluctant to generate GP budgets based on it. Further work is warranted on why there may be mixtures of distributions in the dataset, similar to the work done by Deb and Trivedi (1997, 2002) on a 'healthy' / 'ill' dichotomy in health service utilisation.

By the same reasoning, we are reluctant to apply iteratively re-weighted least squares in the absence of a theory of why some observations are outliers and others are not, especially when it has such a dramatic effect on results.

If we are to truncate, we face a choice about the truncation point. The current scheme assigns certain drugs as budget neutral, excluding them from budgets. It also sets a truncation point at IR£2075, which is at the 99.1st percentile of the dataset. It seems reasonable that the budget neutral provision is maintained. However, the truncation point is debatable. The exploratory analysis offers a number of options.

First, the maximum value for observations with a non-zero weight using iteratively re-weighted least squares and excluding budget neutral drugs is IR£1,170.90 which is at the 95.9th percentile. Although this is just 4.1% of observations it would exclude 19% of total expenditure. Second, quantile regression at the 95th quantile produces markedly different results to those at other quantiles, suggesting that a truncation at about this quantile should be considered. The 95th percentile of the distribution excluding budget neutral drugs is IR£935.14,which excludes 23% of all expenditure. Third, the two-component finite mixture model indicates that the 97th quantile should be considered, which is IR£1,315.80. This would exclude 17% of total expenditure. Fourth, the analysis using Cooks D finds that the maximum value of the non-outliers is IR£1,719.35, when budget neutral drugs are excluded, which is at the 98.6th percentile, meaning that 13% of total expenditure would be excluded.

Fifth, the current IDTS includes a provision for high cost patients, which were defined as patients costing more than IR£2,075 in 2000, meaning that all further expenditures for these patients were assumed by the GMS (Payments) Board, not the GP. This is the 99.1st percentile for our dataset. At this truncation point, 11% of total expenditure is excluded. Sixth, many risk-adjustment studies allow for severe outliers by top-slicing the dataset at the 99.9th percentile (Deb and Burgess, 2002). This would exclude just 1% of total expenditure.

In conclusion we favour truncation as a means of accounting for 'outliers'. Since most studies in the literature apply minute levels of truncation, and the precedent in Ireland is for truncation at the 99.1st percentile we wish to be conservative. On the principle that we do not want to assign an observation as an outlier that may not be an outlier, and given that the maximum value for non-outliers using Cooks D is IR£1719.35, we use this as our cut-off value.

As discussed in Chapter 2, Newhouse (1996, 1998) points out that given the inability of most risk-adjustment studies to explain any more than about 20% of variance, the budget-holder bears the majority (and sometimes all) of the risk associated with high expenditure patients, for which they are not adequately compensated. Newhouse (1998) recommends partial capitation, also known as 'supply-side cost sharing' to overcome this problem. However, this broadens the discussion from the effect of

outliers on the estimation of a utilisation function to their effect on the risk exposure of the budget holder. If all GPs had their fair share of outliers, then their risk exposure due to outliers would not be a cause of concern. To the extent that some GPs are more exposed to budget deficits simply because they are unlucky enough to have a disproportionate number of outliers – meaning high cost patients - on their panel, they are a cause of concern for us. Therefore an integral part of the decision of what to do with high cost patients is an assessment of the risk borne by GPs at different levels of truncation. While we postpone this analysis to Chapter 7, we can say that the results are qualitatively similar to those outlined here – the greater the extent of truncation the lower the risk to the GP but no level of truncation stands out as a clear favourite.

One caveat is in order. This analysis has focused on the additive linear model, as it is clearly the most popular model applied to risk-adjustment. However, it is possible that a proportional model may be a better representation of the relationship between the set of covariates and prescribing expenditure. Those observations that are considered outliers in an additive setting may be non-outliers under a transformed model. The next chapter compares the additive linear model, specified as OLS and truncated at the 98.6^{th} percentile, with a number of transformed models with the same level of truncation. Appendix 5.1 examines results at different levels of truncation.

4.7 CONCLUSIONS

This empirical assessment of outstanding issues in model specification highlighted a number of important results. First, a review of the economics literature found a number of different specifications of age were used in econometric models similar to the one applied here. Moreover, the quadratic specification has been heavily criticised (Murphy and Welch, 1990). We found, however, that we could not distinguish statistically between four specifications and chose the quadratic one on the grounds of popularity and its relative lack of arbitrariness.

Section 4.2 demonstrated the extent of the skew in the distribution of prescribing expenditure. The next three sections were devoted to ways of analysing skewed prescribing expenditure data. We found that previous risk-adjustment studies had

removed 'outliers', but had identified them in relatively ad hoc ways. Section 4.3 outlined and applied rigorous outlier identification techniques, and found that the number of 'outliers' was considerable. The majority of these were high cost patients. Robust regression also found that observations that were assigned a weight of zero were much more likely to be indicated as having chronic diseases, a disability and be older than other observations.

Section 4.4 described the relationship between covariates and the response variable at different points on the conditional distribution, using quantile regression. This pointed to considerable heterogeneity in the slopes of some covariates at different points on the conditional distribution, most obviously the chronic disease indicators, age and age squares and disability. It also emphasised the difference between high cost patients and the rest of the sample, with a clear kink from the 80^{th} quantile to the 95^{th} quantile.

Section 4.5 reported that finite mixture models had been used successfully for risk-adjustment in two cases. We tested their usefulness in the GMS setting. We found that spurious results were generated for many classes, so the application of the approach was not justified. We recommend further research into the possible causes of the mixtures of distributions that were identified. In addition, this modelling exercise again pointed to differences between high cost patients and the rest of the sample.

Given the clear evidence that high cost patients are different from the rest of the sample, section 4.6 considered ways of dealing with them. It ruled out applying iteratively-reweighted least squares and finite mixture models as they lacked a theoretical basis, and favoured the truncation of the dataset. The truncation point was chosen as the maximum value of non-outliers based on Cooks D, once budget neutral drugs were excluded, which was found to be IR£1719.35.

5. EMPIRICAL ESTIMATION OF UTILISATION MODEL

As recommended by Schokkaert and van de Voorde (2000), we split the risk-adjustment process into two phases – a so-called 'empirical' phase that explains the determinants of prescribing expenditure and a so-called 'normative' phase that generates budgets. This chapter is concerned with the empirical phase, while subsequent chapters focus on the normative phase. We explain the determinants of prescribing in the context, not only of other risk-adjustment studies, but also of other studies of health care utilisation functions. We examine variables for counter-intuitive signs, drawing on the review of theoretical and empirical models in Chapter 2. Any such variables are then treated differently in setting budgets in subsequent chapters.

In addition, there is a paucity of research on the health care utilisation in Ireland, so an explanation of the determinants of prescribing expenditure in Ireland is welcome. Therefore, we attempt to contribute to the general health econometric literature, and especially to comment on those variables that were identified as having ambiguous signs in Chapter 2, as well as to the understanding of Irish health care utilisation.

A number of empirical models of prescribing expenditure were proposed in Chapter 2. We referred to them as the Demographic model, which consisted of demographic and socio-economic variables and is similar to the so-called Demographic model applied in Dutch risk-adjustment studies; the Chronic Illness model, which was the Demographic model supplemented with measures of chronic illness and finally the Supply model, which is the Chronic Illness model supplemented with supply-side variables. These models are nested, so on statistical grounds we should test the joint significance of each additional group of variables. If they are significant, then we should choose that model. However, we argued above that because chronic illness may be subject to measurement error related to prescribing style of the GP and certain supply-side variables have not been used before and therefore should be applied with caution. Therefore we report the results of all three models here. This approach also illustrates the effect of additional variables on explained variance and parameters.

We also outlined a number of competing estimators in Chapter 2. While OLS is by far the most popular estimator used in risk-adjustment, the two-part model is a feasible alternative. This can potentially overcomes problems with a mass of observations at zero, a right skew and a heavy right tail, as were highlighted in Chapter 4. Both one-part and two-part models are estimated here.

Section 5.1 outlines the one- and two-part models estimated in this chapter. Results of one-part models are described in section 5.2, while section 5.3 describes results for two-part models. These sets of results are drawn together and discussed in section 5.4, while section 5.5 offers conclusions.

5.1 METHODS

One-part models are specified as follows:

$$y_{ij} = \beta x_{ij} + \phi z_j + u_{ij}, \qquad (5.1)$$

where y_{ij} is prescribing expenditure of individual i attending GP j, x_{ij} is a set of individual-level covariates, z_j is a set of GP-level covariates, u_{ij} is an i.i.d. error term and β and ϕ are parameter estimates. For the Demographic and Chronic Illness models, the z_j vector is absent. The model is estimated using OLS.

The first part of the two-part model are specified as either a probit model or a logistic regression. We choose the former. More formally, the first part is specified as

$\text{Prob}[y_{ij} = 1] = \Phi(\alpha x_{ij}, \varphi z_j)$,

where y_{ij} is one if the individual has positive prescribing expenditure and zero if the individual has zero prescribing expenditure, α and φ represent the set of parameters to be estimated and $\Phi(.)$ is the standard normal distribution.

There is no clear consensus on the correct estimator to apply in the second part of the two-part model. The only way to choose between these competing estimators is through empirical application. The first specification we consider is a log-transformation, specified as follows:

$\ln(y_{ij} / y_{ij}>0) = \delta x_{ij} + \theta z_j + e_{ij}$

where δ and θ represent a set of parameters to be estimated and e_{ij} is an i.i.d error term.

The second estimator we apply is based on the GLM, (Blough et al., 1999). We briefly summarise the description of the GLM in Chapter 2 here. There are three parts to the GLM. First there is the linear predictor, similar to OLS:

$\eta_{ij} = \gamma x_{ij} + \lambda z_j$.

Second, there is a monotonic differentiable link function which maps the linear predictor onto the expectation of the response variable. An array of link functions can be specified, although the logarithmic link is popular for health expenditures (Blough et al., 1999; Madden et al., 2000; Manning and Mullahy, 2001):

$\ln(\mu_{ij}) = \gamma x_{ij} + \lambda z_j$,

where $\mu_{ij} = E(y_{ij})$. Third, there is a variance function describing how the variance depends on the mean. We specify the variance function as a gamma distribution, as health utilisation data are often characterised by variance approximately equalling the square of the mean (Blough et al., 1999):

$Var(y_{ij}) = \sigma_{ij}^2 = \kappa(\mu_{ij})^2$,

where κ is a constant called the dispersion parameter.

The interpretation of coefficients from both these models is straightforward. If covariate h is specified in natural logarithms, then the elasticity of utilisation with respect to h is:

$\varepsilon_h = (1-P)\alpha_h + \beta_h$,

for the logarithmic model, where P is the probability of getting a prescription, while if covariate k is specified in natural units, elasticity is measured as:

$\varepsilon_k = [(1-P)\alpha_k + \beta_k]x_k$.

These formulae assume that the error term is homoscedastic in covariate h or k.

We compare the goodness of fit of models using the generalised R^2 (Madden et al., 2000) as follows:

$$R^2 = 1 - \exp\{-\frac{2}{N}\sum_{m=1}^{2}(l(\theta_m) - l(0_m))\},$$

where $l(\theta_m)$ and $l(0_m)$ are the log likelihoods of the fitted model and the null models respectively. This measures the percentage of explained variation and allows comparison across all models, including GLM.

As with all models estimated in the study, cluster robust standard errors are reported, which provides heteroscedasticity robust standard errors that have the additional property that the assumption of independence of observations within a GP practice is relaxed.

5.2 RESULTS

Table 5.1 describes the determinants of prescribing expenditure as modelled using OLS.

Table 5.1

Determinants of Prescribing Expenditure using One-Part Models

Variables	NARA		Demographic		Chronic Illness		Supply	
	Coef.	P>|t|	Coef.	P>|t|	Coef.	P>|t|	Coef.	P>|t|
constant	249.16	0.00	35.40	0.00	10.11	0.00	-42.10	0.13
age04	-219.53	0.00						
age515	-223.87	0.00						
age1644	-144.40	0.00						
age6569	76.92	0.00						
age70	95.84	0.00						
Age			3.56	0.00	2.81	0.00	2.81	0.00
Agesq			0.01	0.00	-0.01	0.00	-0.01	0.00
Gender			-1.38	0.17	-1.12	0.24	-1.22	0.24
Marital			16.67	0.00	16.03	0.00	16.05	0.00
Disabil			100.22	0.00	49.59	0.00	47.21	0.00
Lonepare			-0.45	0.87	3.59	0.10	3.48	0.13
Numcard			-14.81	0.00	-8.90	0.00	-8.77	0.00
Rural			-11.15	0.00	-7.11	0.00	0.09	0.94
3-5 miles			-7.87	0.00	-4.17	0.00	-3.91	0.01
5-7 miles			-9.18	0.00	-7.65	0.00	-6.85	0.00
7-10 miles			-25.35	0.00	-15.99	0.00	-12.56	0.00
10+ miles			-22.88	0.00	-16.20	0.00	-10.08	0.02
NEHB			6.29	0.00	21.55	0.00	22.11	0.00
SEHB			6.64	0.00	10.78	0.00	16.07	0.00
CVD					261.45	0.00	259.83	0.00
Epi					274.63	0.00	274.79	0.00
Rheum					316.39	0.00	312.19	0.00
Diabetes					383.20	0.00	380.44	0.00
Glau					270.93	0.00	269.69	0.00
Respir					331.69	0.00	328.84	0.00
Thyroid					107.09	0.00	107.32	0.00
Psych					354.93	0.00	352.62	0.00
comor					-64.06	0.00	-62.50	0.00
GPage							-6.64	0.00
Gpagesq							0.06	0.00
Nurse							8.92	0.00
Sec							-4.38	0.02
RPA							-5.22	0.00
Decpanel							0.00	0.03
Specific							149.73	0.00

symptomatic							427.74	0.00
Presum							348.51	0.00
N	400,751		400,751		362,072		305,887	
RMSE	315.3		312.82		269.70		269.71	
$\overline{R}^2$ (%)		13.23		14.59		38.75		38.91

All four models fail the Ramsey Reset test. This is discussed below. The $\overline{R}^2$ for the NARA at 13.23% is extremely high. Most risk-adjustment studies find that age and gender explain between 1% and 3% of variance (Newhouse et al., 1989). While we truncated the response variable much lower than most risk-adjustment studies do, the $\overline{R}^2$ when all expenditure is included is approximately 10%, which is still extremely high. We suggest that since GPs have had their budgets set on the basis largely of the NARA since 1994, this has affected their prescribing habits to the extent that the NARA now explains far more variance in prescribing than would age and gender if GPs faced no budget constraint. This is also discussed further below.

The reference class for the NARA is people aged between 45 and 64. The coefficients in the NARA model are as expected. People aged 45 to 64 have annual prescribing expenditure of approximately IR£250. Younger age groups have less expenditure, with those aged 5 to 15 having the lowest expenditures, and older age groups having higher expenditure. The over 70s have average prescribing expenditure of IR£345 according to the NARA model. These results are consistent with those presented in section 4.1. However, the patterns exhibited in Figure 4.1 suggest that the NARA age bands are too wide. For instance, the reduction in expenditure in those over 80 is not captured by the NARA model.

Next, we consider the Demographic, Chronic Illness and Supply models. Following multiple imputation, there is a full set of observations for the Demographic model. The removal of the 38,447 individuals who were not on the GMS scheme during the Chronic Illness measurement period reduces the sample size of the Chronic Illness model to 362,072, while the removal of individuals attending those GPs for whom we do not have reliable indicators reduces the sample size for the Supply model to 305,887. In unreported work, we restricted analysis to the 305,887 observations included in the Supply model. The coefficients on the Demographic model and Chronic Illness model for this sub-sample were very similar to those reported here.

Therefore, despite the fact that the models are estimated on different samples, it is reasonable to compare coefficients across them. In addition, the $\overline{R}^2$ of the Chronic Illness was also 38.75% when the sample size was restricted to 305,887, illustrating the negligible marginal effect of sample size on explained variance in samples as large as these.

The $\overline{R}^2$ of the Demographic model is 14.59, only slightly higher than that of the NARA model. Meanwhile the $\overline{R}^2$ of the Chronic Illness and Supply models are very similar at 38.75 and 38.91 respectively. These are again high for risk-adjustment studies. Breyer (2001) has an $\overline{R}^2$ of 37%, but the model includes past expenditures as a covariate, which increases explained variance considerably. Madden et al. (2000) produces an $\overline{R}^2$ of 51%, but they use a two-part GLM model. Most one-part OLS models that include health status as a set of covariates produce $\overline{R}^2$s of the order of 11% to 14% (Lamers, 1999a; Lamers and van Vliet, 2001). However, as outlined above, age explains much more variance in this model than is does in most others, so this pushes up the explained variance for our other models as well.

As a general pattern, coefficients in all three models are signed with expectations, while the difference between the Chronic Illness model and the Supply model is very modest for most coefficients. It appears that the effect of including supply-side variables is slight, although supply-side variables are jointly significant according to an F-test.

The reference categories for the distance and Health Board dummy variables are less than three miles from the GP's principal surgery and the Southern Health Board respectively, as these best approximated the mean prescribing expenditure. Expenditure is quadratic in age for all three models. However, for the Demographic model is squared term is slightly positive, while it is slightly negative, as expected, for the Chronic Illness and Supply models. In fact, the negative squared terms are so slight that the series does not reach a maximum over the age range of the sample. The combined effect of the age terms for the Demographic model is that a 45 year old has an expected expenditure of IR£171.28 greater than a one-year old; the expected difference is IR£93.21 for the Chronic Illness model and IR£94.39 for the

Supply model. Hence, the introduction of chronic illness variables reduces the importance of age.

After age, disability has the strongest effect on expenditure in the Demographic model. Its effect is approximately halved in the Chronic Illness and Supply models, illustrating the moderating of its effect by the introduction of chronic illness indicators. Marital status is also positively related to expenditure in all three models and by the same amount.

In all three models, gender is insignificant, as is lone parenthood.

The number of people on each medical card is negatively associated with expenditure. We included it as an indicator of poverty, so this may have a counter-intuitive sign. This is discussed below. Those variables that are included to detect access barriers – rural residence and distance to GP – are significant and signed with theoretical expectations, while the pattern of increasingly negative effects on most of the distance indicators is also consistent with expectations. The exception is rural residence in the Supply model, where the introduction of supply-side effects makes rural residence insignificant. Of the supply-side variables, rural residence has the highest partial correlation coefficient with RPA. Both variables appear to be detecting the same phenomenon, which is a negative relationship between rurality and expenditure.

Both residence in the North Eastern Health Board and the South Eastern Health Board are positively associated with expenditure, especially following the introduction of the chronic illness indicators. It appears that expenditure on chronic illnesses is higher in these two Health Boards than in the Southern Health Board, of the order of IR£10 to IR£20 per person.

Chronic illness indicators are very strong determinants of prescribing expenditure. Their effects in rank order are diabetes, psychiatric illness, respiratory illness, rheumatological illness, epilepsy, glaucoma, cardio-vascular disease and thyroid disease. The relatively low coefficient on cardio-vascular disease is partly due to the exclusion of statins as they are budget neutral. This rank order is the same for the

Chronic Illness model and the Supply model. The comorbidity indicator is negative in both models, suggesting that each additional chronic illness adds a little less than the coefficient on that chronic illness. For instance, someone with rheumatology and psychiatric illness is expected to have additional expenditure of IR£607.26 (rheum + psych - comor), not IR£671.32 (rheum + psych), according to the Chronic Illness model.

Supply-side variables are all significant. The effect of GP age is quadratic, negative for all GP ages and quite considerable. Attending a GP who is 50 is associated with IR£186.80 lower expenditure compared to a GP aged 30, a dramatic difference.

Other supply-side effects are very modest by comparison. Attending a GP with a secretary or a GP in receipt of rural practice allowances are associated with slightly lower expenditures, while attending a GP with a practice nurse is associated with slightly higher expenditures. The prescribing style coefficients are relatively modest as well. Attending a GP whose prescribing of specific drugs as a percentage of their total prescribing is 1% above average is associated with a IR£1.50 increase in expenditure, for instance.

Table 5.2 examines the determinants of prescribing expenditure using the two-part modelling strategy. The Supply model is reported, as the supply-side variables were found to be jointly significant in the probit model as well as the logarithmic model and the GLM.

Table 5.2

Determinants of Prescribing Expenditure using Two-Part Models

Variables	Probit		Ln-expend		GLM	
	Coef.	P>|z|	Coef.	P>|t|	Coef.	P>|z|
constant	-3.11	0.00	2.064	0.00	3.978	0.00
Age	0.02	0.00	0.041	0.00	0.044	0.00
Agesq	-0.00	0.00	-0.000	0.00	-0.000	0.00
Gender	0.21	0.00	0.093	0.00	-0.015	0.09
Marital	0.16	0.00	0.061	0.05	0.014	0.31
Disabil	0.10	0.00	0.215	0.00	0.233	0.00
Lonepare	0.24	0.00	0.032	0.00	-0.126	0.00
Rural	-0.03	0.04	0.010	0.17	-0.009	0.52
Numcard	-0.03	0.00	-0.079	0.00	-0.100	0.00
3-5 miles	-0.07	0.00	-0.041	0.00	-0.002	0.88
5-7 miles	-0.07	0.00	-0.043	0.00	-0.024	0.06
7-10 miles	-0.11	0.00	-0.069	0.00	-0.027	0.30
10+ miles	-0.08	0.07	-0.070	0.00	-0.055	0.15
NEHB	-0.01	0.71	0.071	0.00	0.085	0.00
SEHB	0.12	0.00	0.024	0.01	0.033	0.08
CVD	1.69	0.00	1.365	0.00	0.828	0.00
Epi	1.71	0.00	1.533	0.00	1.284	0.00
Rheum	1.53	0.00	1.551	0.00	1.147	0.00
Diabetes	1.61	0.00	1.783	0.00	1.344	0.00
Glau	1.71	0.00	1.508	0.00	0.975	0.00
Respir	1.53	0.00	1.710	0.00	1.247	0.00
Thyroid	1.61	0.00	0.985	0.00	0.689	0.00
Psych	1.51	0.00	1.665	0.00	1.244	0.00
comor	-1.64	0.00	-1.065	0.00	-0.759	0.00
GPage	-0.01	0.65	-0.045	0.00	-0.032	0.00
Gpagesq	0.00	0.72	0.000	0.00	0.000	0.00
Nurse	0.07	0.00	0.054	0.00	0.056	0.01
Sec	-0.04	0.19	-0.040	0.00	-0.069	0.02
RPA	-0.21	0.00	-0.043	0.00	-0.007	0.83
Decpanel	0.00	0.21	0.000	0.03	0.000	0.19
Specific	3.38	0.00	1.118	0.00	0.057	0.89
Symptomatic	4.07	0.00	2.817	0.00	1.311	0.01
Presum	5.61	0.00	3.180	0.00	1.000	0.06
N	305,887		236,308		236,308	
R^2 (%)	14.94		47.24		31.17	

While the Supply model relates to 305,887 individuals, we find here that 236,308 (77%) of these had positive prescribing expenditures. As with the one-part models

above, all three models outlined in Table 5.2 fail the Ramsey Reset test. The χ^2 statistic indicates that the probit model is significant while the pseudo-R^2 is 14.94%, a level similar to many other reported in the prescribing literature. For instance Street et al. (1999) explained between 8% and 13% of the variance in prescribing expenditure in three Russian regions, while Grootendorst (1995) found pseudo-R^2 values of between 22% and 27% for prescribing in Canada, using two-part count models.

The $\overline{R}^2$ for the logarithmic model and GLM are 47.24% and 31.17% respectively, which is high for this type of model. Street et al. (1999) got $\overline{R}^2$s of approximately 10% for a logarithmic model, for instance. On the other hand, Madden et al. (2000) got an $\overline{R}^2$ of between 38% and 51% for the same GLM model that we apply, so our results, while high, are not unrealistic. As explained above, we expect that explained variance is high as a result of the response of GPs to their budgets over time.

Both the probability of getting a prescription and the level of expenditure is quadratic in age, with a positive linear effect and a very slight negative squared term. The combined age elasticity at age 45 is 2.1 for the logarithmic model, versus 0.1 at age 1. In other words a 1 year increase in age at age 45 is expected to lead to a 4% increase in expenditure, while a 1 year increase at age 1 is expected to lead to a 10% increase in expenditure. Age elasticities for the GLM are almost identical. These are plausible magnitudes of effect and are similar to the one-part models.

All other demographic and socio-economic variables are significantly related to probability of use, except living more than 10 miles from the GP and residing in the North Eastern Health Board. Those that are positively related to probability of use are, in rank order, lone parenthood, gender, marital status and disability. Lone parenthood and being female have approximately a 5% higher probability of getting a prescription at the mean of both variables, while the disability variable is associated with a 2% increase in probability at its mean. The combined effects for the logarithmic model are 8.7% for lone parenthood, while it is –0.07% for the GLM.

The combined effect for gender is 14.1% for the logarithmic model, while it is insignificant for the GLM. Disability meanwhile is consistent across models,

strongly affecting probability and extent of use. The combined effect of disability for the logarithmic model is 23.8% and 25.6% for the GLM. Marital status is also reasonably consistent. It has a positive effect on probability of use and for the logarithmic model. It is, however, insignificant in the GLM. The number of people on each medical card is always negative. It is associated with a 0.7% decrease in probability of use at its mean, while the combined effect for the logarithmic model and GLM is a decrease of 8.6% and 10.7% respectively.

The effect of rural residence and distance to the GP on probability of use is negative as expected. Rural residence is insignificant in both the logarithmic model and GLM and all distance dummies are insignificant in the GLM. Our interpretation of the GLM specification is that these variables affect the probability of accessing the GP but not the level of expenditure once the GP is accessed. This is as expected. They do, however, have significant negative effects in the logarithmic model, suggesting that they affect the level of expenditure as well the probability of attendance in this model.

Chronic illness indicators all strongly affect the probability of getting a prescription. At their means, glaucoma is associated with a 16% increase in probability of attendance and cardio-vascular disease is associated with a 23% increase. The other chronic illnesses vary between these two. They also have large positive coefficients on both the logarithmic model and the GLM. For the logarithmic model, thyroid disease is associated with a 135% increase in expenditure in total and diabetes is associated with a 215% increase. Other indicators are in between these extremities. The combined effects in the GLM are slightly less, varying from an increase of 105% for thyroid disease to 171% for diabetes. As with the one-part model, comorbidities are negatively associated with probability of use and level of use for both models. The combined effect for the logarithmic model and GLM is 143% and 113% respectively for each additional comorbidity. This means that for most chronic illnesses, having comorbidities is associated with increased expenditure, but not increases as large as the combined effect of the coefficients on the multiple chronic illness indicators.

The supply-side variables are not as important as other variables in determining the probability of use. This is consistent with the idea that patients are the principal decision-makers in the first part of the two-part model and physicians are more important in the second part. We find that GP age, having a secretary and panel size are insignificant. Having a nurse is associated with a 2% increase in probability of use at the mean, while attending a GP in receipt of rural practice allowances is associated with a 6% reduction in probability of use at the mean, as this variable is included to detect access barriers. The GP prescribing style variables are all positive but their effect is slight.

The GP age effect is significant in both the logarithmic model and GLM, however, meaning that the combined effect for both these models is –4.6% and –3.3% respectively. The combined effect of the nurse variable is 7.1% and 7.3%, while the same results for secretary are –4.8% and –7.7%. Rural practice allowance is associated with a total decrease of –9% for the logarithmic model and it is insignificant for the GLM. Finally the prescribing style variables have significant positive total effects in the logarithmic model. For instance, a 1% increase in the percentage of total prescribing that is often presumptive is associated with a 4.5% increase in expenditure. For the GLM, only the symptomatic variable is significant.

5.3 DISCUSSION

All models fail the Ramsey Reset test. This is a general test for misspecification, which adds the squares and higher polynomials of the fitted values as additional regressors and tests their significance. Given the size of our dataset, it is unsurprising that such a test detects significant effects. In order to check, we took a random sample of 1000, applied a Ramsey Reset test and repeated the procedure 20 times. It passed on each occasion.

The Reset test can either relate to misspecification of functional form or to omitted variables. When we consider that not only have we applied three different estimators in this chapter, but also examined variations on the specification of age in Chapter 4, then most of the candidate functional forms have been applied, yet they all still fail the test. If there are omitted variables, then to the extent that they are uncorrelated with the current set of variables, the model will be inefficient but not inconsistent.

Given the size of the dataset, levels of inefficiency would have to be considerable before the results would be called into question. If omitted variables are correlated with the current set of regressors, then the model will be inconsistent also. However, we have included more variables than most other risk-adjustment studies, so we should have much lower omitted variable bias. Therefore, given the sample size used, the fact that all models fail the Reset test should not be too great a cause of concern.

We suggested that the high $\overline{R}^2$ on all models is due to the unobserved response of GPs to their budget, rather than the direct effect of age on expenditure. The finding is interesting as it provides tentative evidence of the effect of the scheme on GP prescribing behaviour. Future research in this area would be interesting. For instance, if our hypothesis is correct, then explained variance should increase year-on-year since the year preceding the scheme's inception, as GPs adjusted their prescribing behaviour in response to their budget. This is a testable hypothesis.

If this conclusion is correct, it has a direct implication on the interpretation of our results for budget setting. Suppose we wish to set prescribing budgets based on need. We assume that, controlling for supply-side effects and access costs, a utilisation function measures the relationship between 'need' and use. However GPs respond to budgets by altering prescribing behaviour, then this effect needs to be controlled to get the independent effect of need on use. The Supply model used here does not include a variable that measures the degree to which a GP alters their prescribing behaviour in response to their prescribing budget. This is a drawback to the 'empirical' approach as opposed to the 'normative' approach discusses in Chapter 1. However, checks are possible. In Chapters 6 and 7, we focus not only on the ability of a risk-adjustment model to obtain unbiased predictions, but on the distribution of those predictions, ensuring that policy-relevant sub-groups are treated 'fairly'. This disadvantage with the 'empirical' approach is well known (Carr-Hill et al., 1994; Sutton and Lock, 2000) but it is still by far the most commonly used approach to weighted capitation modelling, suggesting that most analysts and policy-makers believe that its advantages greatly outweigh its disadvantages. Moreover, to the extent that the payments system ends up compensating GPs for a mix of need and

GP prescribing behaviour, then it becomes a mixture of retrospective and prospective payments, which has its merits (Newhouse, 1998).

We can place the results in Table 5.1 and 5.2 into a number of categories. First, there are variables for which the results are consistent across all models. These include the chronic illness indicators, disability, the number of people on each medical card and the nurse variable.

The chronic illness indicators produce effects similar to previous research. Blough et al. (1999) found that coefficients on the 52 ACGs they included in their risk-adjustment model varied from 0.11 to 5.93 in an extended quasi-likelihood model with a logarithmic link function, similar to the GLM applied here. By comparison, the results in Table 5.2 on log of chronic illness vary from 0.69 for thyroid in the GLM to 1.78 for diabetes in the logarithmic model. Lamers (1999a) and Fishman and Shay (1999) also found large coefficients on their chronic disease score variables, although they are not directly comparable with those in this study as they use different currencies and relate to different care programmes. It is now well established that health status variables derived from diagnostic data or prescribing data dominate explanations of health care expenditure in risk-adjustment studies and this study is consistent with that consensus. We have, however, quantified their effects in an Irish setting for the first time.

In the one-part Demographic model disability is associated with an increase in expenditure of IR£100. Coming from a household in receipt of disability payments is associated with an increase in expenditure of 62% for the average person in the sample. Meanwhile, van de Ven and Ellis (2000) and Breyer (2001) found that disability was associated with an approximate doubling of expenditure. However their results relate only to those individuals who are in receipt of disability payments, while ours relate to individuals who come from households in receipt of disability payments, not all of whom will be in disabled.

While previous research on prescribing expenditure that has included a nurse variable (Whynes et al., 1997) found it to be insignificant, while we find that it is

positively associated with expenditure. Perhaps Whynes et al. (1997) did not have a sufficient sample size to detect a nurse effect.

The number of people on each medical card was included as a potential indicator of poverty, with an expected positive relationship with expenditure. However, we suggested that it may describe unobserved access barriers, such as members of larger families being unable to attend the GP because of child care commitments. Thus we acknowledged in section 2.2.6 that the expected sign was ambiguous. We find, in fact, that it is negatively signed. If the second hypothesis held, then we would expect a comparatively large negative effect in the first part of the two-part model. In fact, the variable has a small effect on probability of use but a much larger effect on the extent of use given any use. Examining the one-part models, it is noteworthy that the coefficient drops from the Demographic model to the Chronic Illness model. Those with chronic illnesses are from smaller households than the average, so the numcard variable could be representing a phenomenon that unhealthier people appear to come from smaller households. Numcard is also negatively related to age, so it might be detecting some part of the lower expenditure of younger people that the age specification is failing to detect. Finally, it may be detecting the ability of a family to care for someone who if living alone would need to attend to GP and may therefore get a prescription. In conclusion, rather than detecting poverty or access barriers, the variable appears to be detecting good health or at least a reduction in the need for prescribing. We suggest that the negative sign is not counter-intuitive. Further research should examine the variable further. For instance, rather than generate a variable relating to number in the household only, we could identify those who live alone and check if this has an independent positive effect on health care need. In addition, it would be interesting to test its effect on particular chronic illnesses such as psychiatric illness.

Second, there are a number of variables that have consistent results and are insignificant in the first part of the two-part model. These include the GP age specification, the secretary variable and health board of residence (which are jointly insignificant in the probit model and jointly significant in all others). We are not surprised that these variables are insignificant in the first part of the two-part model

because they are supply-side effects which are less likely to affect the contact decision, so the pattern is consistent with expectations.

The overall results on the GP age specification is of interest. Although other studies have found GP age (or experience) to be significant predictors of utilisation (Sorensen and Grytten, 1999; Wilensky and Rossiter, 1981) or fees (Goldman and Grossman, 1978; Gaynor and Polachek, 1995), they never found it to have the magnitude of effect detected here, while other studies have found physician age to be insignificant (Gravelle et al., 2002; Sorensen and Grytten, 2001). Since GP age is not subject to measurement error, we take the result at face value. The reason why GP age would have such a large effect in this study and not others, coupled with the result in Chapter 4 that GP age differed by quantile with older GPs treating high expenditure patients different to younger GPs, suggests that this variable warrants further study.

Third, there are variables for which results are universally consistent but for one exception. These include the age specification, distance indicators (when considering their joint significance), marital status, rural practice allowance, and the prescribing style indicators. The GLM is the exception for all of these except the age specification, which is different in the Demographic model.

The quadratic specification in age is as expected for most models, with a large positive effect for the first order polynomial and a slight negative effect for the second order polynomial. This is consistent with exploratory analysis in Chapter 4 and previous research as discussed in Chapter 4. The one exception is the Demographic model where the second order polynomial is positive.

This is one of the few individual-level health care utilisation functions that we identified that includes distance as a covariate, the others being Geil et al. (1997) and Ryan et al. (1999). Given that we expect price to affect utilisation and in a health care system free at the point of contact, time prices are likely to be important, so this is a gap in the literature. We find distance is a negative covariate of prescribing expenditure. According to the two-part logarithmic model, living more than 10 miles from the GP is associated with a 8.8% decrease in use, while the percentage effects

for one-part models are of similar orders of magnitude. Although not comparable directly with any previous literature, we note that Haynes et al. (1999) found that each additional kilometre from the GP surgery was associated with a 0.95% reduction in acute admissions, whereas if we average the result for those living more than 10 miles from the GP, then we find that each additional kilometre is associated with a 0.55% reduction in prescribing expenditure. Meanwhile, the rural residence indicators and the rural practice allowance indicator also represent access difficulties such that an individual who lives more than 10 miles from the GP, is indicated as a rural dweller and attends a GP in receipt of rural practice allowance has 17.9% lower expenditure than the average according to the two-part logarithmic model. Less that 1% of the sample are in this category, however.

We found in Chapter 2 that one study found that being married was positively related to utilisation, while one found it to be insignificant. We find that it is positive, lending support to that contention. However the effect of marital status is not generated from theory, and if it is seen as an important policy issue then further research would be required before a consensus could be achieved. It is, however, interesting to note the results of the income model above found that marital status is associated with a higher probability of being from a household with full GMS eligibility, and was negatively associated with income. Therefore, the marital status result here could be detecting the effect of low income on health care need.

The coefficients on symptomatic and often presumptive are as expected, but we expected a negative coefficient on specific. If GPs prescribe a higher proportion of specific medicine, they prescribe a higher proportion of medicine only after a positive diagnosis had been reached, so they should be more conservative prescribers. However there is a residual category, such that assigning a larger proportion of medicine to specific drugs, does not mean assigning a smaller proportion to symptomatic and often presumptive drugs. A GP could assign a larger proportion to all three and less to the residual, in which case the specific category could be detecting chronic illnesses not included in the chronic illness indicators.

As outlined in Chapter 2, the design and application of new indicators of prescribing style is one of this study's innovations. They are all found to be significant and are

promising measures of physician behaviour, especially in an areas of research that finds physician effects to be considerable (Phelps, 2000) but measures of physician effects to be so often insignificant (Gravelle et al. 2002, for example).

We have no clear understanding of why the above variables are insignificant in the GLM. Further exploration of the dataset is required to understand why distance, marital status, rural practice allowance and prescribing style are all insignificant if the variance function is based on a gamma distribution but significant if it is based on a Gaussian distribution.

That leaves four variables with inconsistent results across models, namely, gender, lone parenthood, rural residence and panel size. Gender and lone parenthood are insignificant in the one-part models and significant in the first part of the two-part models. Gender is also significant in the second part of the logarithmic model, while lone parenthood is significant in both second parts but with opposite signs. Rural residence is negative and significant in the first part of the two-part model and two of the one-part models and insignificant in three others. We suggested in Chapter 3, however, that, as there is no formal definition of rurality, this variable may be subject to idiosyncratic measurement error. Panel size is significant in two models but it effect is tiny.

Inconsistency in the sign on gender is a feature of previous research. While it is generally positive, Watson (1996) found that gender was negatively associated with GP utilisation for Ireland. Thus its inconsistency here is not unexpected.

Although lone parenthood was included as an indicator of poverty, and therefore was expected to be positively associated with expenditure, our analysis of the Household Budget Survey in Chapter 3 and of prescribing here suggests that the lone parenthood does not necessarily detect poverty. We found that households headed by lone parents were no more likely that others to be fully entitled to GMS services, *ceteris paribus*, nor had this indicator any effect on income. In other words, we find no evidence of this group being more likely to live in poverty. This finding contradicts Callan et al. (1996) who identified lone parents as being high risk of suffering poverty. However, the Irish economy has grown remarkably since Callan et al.'s

(1996) study, and children's allowance has also grown significantly, so lone parents may no longer be high risk of poverty. Further research into this area is recommended.

Bearing in mind the income results reported in Chapter 3, it is not so unexpected that we find an insignificant relationship between the lone parent variable and expenditure in the one-part models. However, the variable has an important effect on probability of use and overall is positive for the logarithmic model and slightly negative for the GLM. An overall pattern is difficult to discern. What is clear, however, is that it serves some purpose other than to indicate poverty.

Therefore all but four variables are reasonably consistent across estimators. Of the four that are inconsistent, rurality is likely to be subject to measurement error, panel size does not have an economically important effect. The majority of variables are in line with theoretical expectations. As a result, no variables are assigned as 'counter-intuitive' in the budget setting exercise in the next chapter.

The results highlight the importance of health status and age on health care utilisation, the negative effect of distance to the GP and, interestingly, the importance of supply-side effects. The first result is extremely important for risk-adjustment formulae. While we focus on the prescribing formula here, GMS capitation payments for GP services do not consider health status in determining GP income. These results suggest that such an exercise is worthy of consideration. The final result finds that not only are supply-side effects jointly significant, they are economically important. Except for the first part of the two-part model where it is insignificant, GP age has an important role. This is also one of the few occasions when prescribing style has been modelled explicitly in a utilisation study. We find that prescribing style variables are significant, economically and statistically and encourage other researchers to experiment with these variables or use these ideas to generate suitable variables of their own.

This analysis is based on a truncated response variable. Appendix 5.1 reports the results when all expenditure is included and when truncation is based on the current IDTS provisions. The results are broadly similar to those presented here.

5.4 CONCLUSIONS

This chapter examined the determinants of prescribing expenditure for recipients of GMS services in Ireland. Six utilisation functions were estimated. Results were largely in line with expectations, and effects were similar to those reported in previous research. The story told by the one-part models is reasonably similar to that told by the two-part models, insofar as the same variables are easily the most important across models.

We found that the level of explained variance was far in excess of that reported in the majority of risk-adjustment studies. We suggested that this was due to the effect of age as a needs variable being contaminated by the fact that the current budget setting formula is based largely on age and so this variable may be detecting GP's response to the formula as well as the effect of age on need. This could be easily tested empirically, which would provide interesting information on physician behaviour, as well as assist us in an assessment of how best to model age in a risk-adjustment formula that has used age previously to set budgets.

The most economically significant covariates were chronic illness, age and disability. Other important covariates were marital status, distance to the GP and supply-side effects, especially GP age and GP practice style. The marital status variable may be detecting low income. We found that the numcard variable appeared to be detected otherwise unobserved good health, although we recommended reconstructing this variable to examine the effect of living alone on prescribing need in future research, as we suspected that this variable was detecting the positive effects of not living alone. In addition, the lone parent variable appears to be counter-intuitive *prima facie*, but we argued that it no longer reflected poverty in Ireland, given the results of the income model in Chapter 3. This is, however, a tentative conclusion and is worthy of further research.

6. MODEL PREDICTIONS AND DISTRIBUTION OF EFFECTS

This chapter examines the consequences of making the step from estimating a well-specified utilisation function to setting budgets. This includes an examination of the prediction of 'need', a thorough assessment of prediction error, and the distribution of prediction error with respect to income and other policy-relevant groups. We wish to choose a model that generates a vector of predictions that are unbiased, precise and fair.

In Chapter 2, we found that the standard risk-adjustment model only includes needs variables as risk-adjusters. The critique of this approach suggested that all variables with a material effect on utilisation should be included in the utilisation equation and only needs variables should be used for generating capitations. We identified distance to GP, rural residence and GP characteristics as variables that affected utilisation but not need and so should be controlled for rather than included in the capitation formula. Meanwhile we found that there were no counter-intuitive variables in the utilisation equation. Therefore, section 6.1 begins this chapter by outlining how to control for non-needs variables in order to predict 'need' for each individual. We also discuss the role of unmet need in this study.

According to Klein (1992) *"the ability to make useful ex-ante forecasts is the real test of a model"*. In order to assess relative prediction performance, as well as measures of explained variance, we compare variation in predictions with actual variation and examine the prediction bias using the Mincer-Zarnowitz test (Mincer and Zarnowitz, 1969). We also measure the distribution of predictions with respect to income using quintile shares and concentration indices. We also examine each model's ability to predict need for vulnerable groups. As such, this study offers a more comprehensive assessment of each model's prediction performance than many other risk-adjustment studies reviewed in Chapter 2, which concentrate chiefly on explained variance. Indeed, we know of no other individual-level risk-adjustment study that examines the income-related distribution of predictions, while the generation of data on vulnerable groups is a useful and novel approach to model evaluation. The Mincer-Zarnowitz test has been applied in a number of US studies

(Fishman and Shay, 1999; Hornbrook et al., 2001) but has never been applied in a European study, as far as we are aware.

In order to conduct prediction analysis, the sample of 400,751 is split randomly and evenly into an estimation sample and a prediction sample. Since the observations used for estimation are not used for prediction, overfitting is prevented. This process is repeated 50 times in order to generate estimates of the variation in each prediction statistic (Hornbrook and Goodman, 1995; Deb and Burgess, 2002). Variation is expressed as the standard error. Where statistical significance is referred to in the text, it is calculated using a paired t-test.

We wish to evaluate the performance of seven models. First there is the NARA. Next there are the three one-part models – the Demographic, Chronic Illness and Supply models. Then there are the two-part models, both of which are estimated on the full set of variables, that is, the Supply model. In Chapter 2 we discussed the retransformation problem of the logarithmic model in the presence of heteroscedastic errors on the logarithmic scale. Therefore, we apply two retransformations of the logarithmic model. This first is the Duan smear (Duan, 1983), which is the sample average of the exponentiated errors, and assumes logarithmic scale homoscedasticity, as outlined in section 2.2.3. The second is the Manning retransformation (Manning, 1998) which is the predicted value of the exponentiated errors as described in equation (2.25), and assumes logarithmic scale heteroscedasticity. Finally, we evaluate the two-part GLM described in Chapter 5.

Section 6.1 examines the prediction of 'need'. Section 6.2 outlines the methods used for prediction analysis, while section 6.3 describes the methods used for distribution analysis. Section 6.4 presents the results of the prediction and distribution analysis and section 6.5 discusses the implications for model selection.

6.1 PREDICTION OF 'NEED'

Most risk-adjustment studies include 'need' related variables only in their utilisation function (see for instance van de Ven and Ellis (2000)). However, we argued in Chapter 2 that all variables that may have a systematic effect on health care utilisation should be included in the econometric model and these should then be

separated into 'legitimate' variables for inclusion in a weighted capitation formula and other 'nuisance' variables to be controlled. Therefore, need is measured as follows:

$$N_i = \sum_{n=1}^{N} x_{ni}\hat{\beta}_n + \sum_{a=1}^{A} \bar{x}_{ai}\hat{\gamma}_a + \sum_{j=1}^{J} \bar{z}_{ij}\hat{\delta}_j \,, \qquad (6.1)$$

where N_i is need prediction for individual i, x_{ni} is the vector of 'legitimate' regressors N for individual i, $\bar{x}_{ai}$ is the mean values of the vector of 'nuisance' regressors a, $\bar{z}_{ij}$ is the mean values of the vector of provides characteristics that the individual attends, and $\hat{B}_j$, $\hat{\gamma}_k$ and $\hat{\delta}_j$ are parameters. Supply-side variables, distance to GP, which measures the price of health care in our model, and rural residence are 'nuisance' variables and these are set to their sample averages in order to calculate an individual's need prediction.

A second consideration in generating predictions of 'need' from the above models is that a full set of regressors is not available for all observations. The Supply model covers 76% of observations, while the Chronic Illness model covers 90% of observations. To generate predictions for all observations, we apply the next most unrestricted model to remaining observations for which that model can generate predictions. This applies to both one-part and two-part models. For instance, predictions for the Supply model are generated as follows:

$$N_{iS} = \begin{cases} N_{iS} & if\ S_i \neq . \\ N_{iCI} & if\ S_i = . \ \&\ CI_i \neq . \\ N_{iD} & if\ S_i = . \ \&\ CI_i = . \end{cases} \qquad (6.2)$$

where N_{iS} is the need prediction in the Supply model, N_{iCI} is the need prediction in the Chronic Illness model, N_{iD} is the need prediction in the Demographic model, S_i are supply-side variables, CI_i are chronic illness variables and D_i are other variables, all for individual i and '.' is a missing value.

Section 2.3 above described how utilisation might be a poor measure of need for health care if certain groups are socially excluded leading to systematic unmet need. Two remedies were proposed. The first involved estimating the relationship between risk-adjusters and utilisation for the sub-sample that are not socially excluded and using these estimates for the entire sample. This was used for a Maori unmet need

adjustment in New Zealand (Rice and Smith, 1999) and as a proposed deprivation-related unmet need adjustment in Scotland (Scottish Executive, 1999). An alternative is to base needs estimates, not on the sample average, but on a more progressive sub-sample. Sutton and Lock (2000) measure the progressivity of health boards in Scotland, and propose applying the needs estimates of the most progressive board to the full sample. Appendix 6.1 tests for unmet need in the GMS and examines the first of these two proposed remedies. It finds that there is little evidence of unmet need and that an unmet need adjustment is unwarranted. Thus we measure need in the remainder of the study as (6.2).

6.2 METHODS FOR PREDICTION ANALYSIS

Based on McCracken and West (2001), Madden (2000) and Hornbrook and Goodman (1995), we outline three tests of prediction performance. First, we wish to minimise prediction error in the validation sample. As well as adjusted-R^2 for the prediction sample, we apply RMSE for the prediction sample and the mean absolute prediction error (MAPE). The RMSE and absolute error statistics measure the average closeness of the predicted value to the observed value, the first penalising especially large errors by squaring the error. Finally, we wish to minimise negative predictions, as they obviously lack face validity. We compute the percentage of total predictions that are negative for each model.

Our second test of prediction performance is variance in predicted values. We prefer models whose prediction variance is similar to actual variance for the validation sample. We include standard deviation and range of predicted values as measures of variance in predicted values.

Biased samples are common in the actual distribution of health care budgets, since a GPs panel is often very different from the national average. As such, the distribution of predictions should be similar to the observed distribution, since models that fit the entire observed distribution are less likely produce errors when applied to biased samples. Therefore, our third prediction test is a test of prediction efficiency and prediction bias. We regress the predicted values in the validation sample on observed expenditures and test that the slope equals unity and intercept equals zero. This is a variant of the Mincer and Zarnowitz (1969) approach, and has been

recommended as good practice by McCracken and West (2001) in a review of prediction analysis. It has been applied in risk-adjustment by Hornbrook and Goodman (1995), and Fishman and Shay (1999).

6.3 METHODS FOR DISTRIBUTION ANALYSIS

One of the objectives of the national health strategy is to ensure equal access for equal need, and that existing variations in health status should be reduced where avoidable. Therefore, on top of the three prediction criteria outlined above, a fourth criterion is that models should be capable of accurately predicting expenditure for policy-relevant vulnerable groups, as described in section 3.6. These include unemployed people, people living alone and those granted discretionary medical cards due to 'hardship'. We examine prediction error for each of these groups.

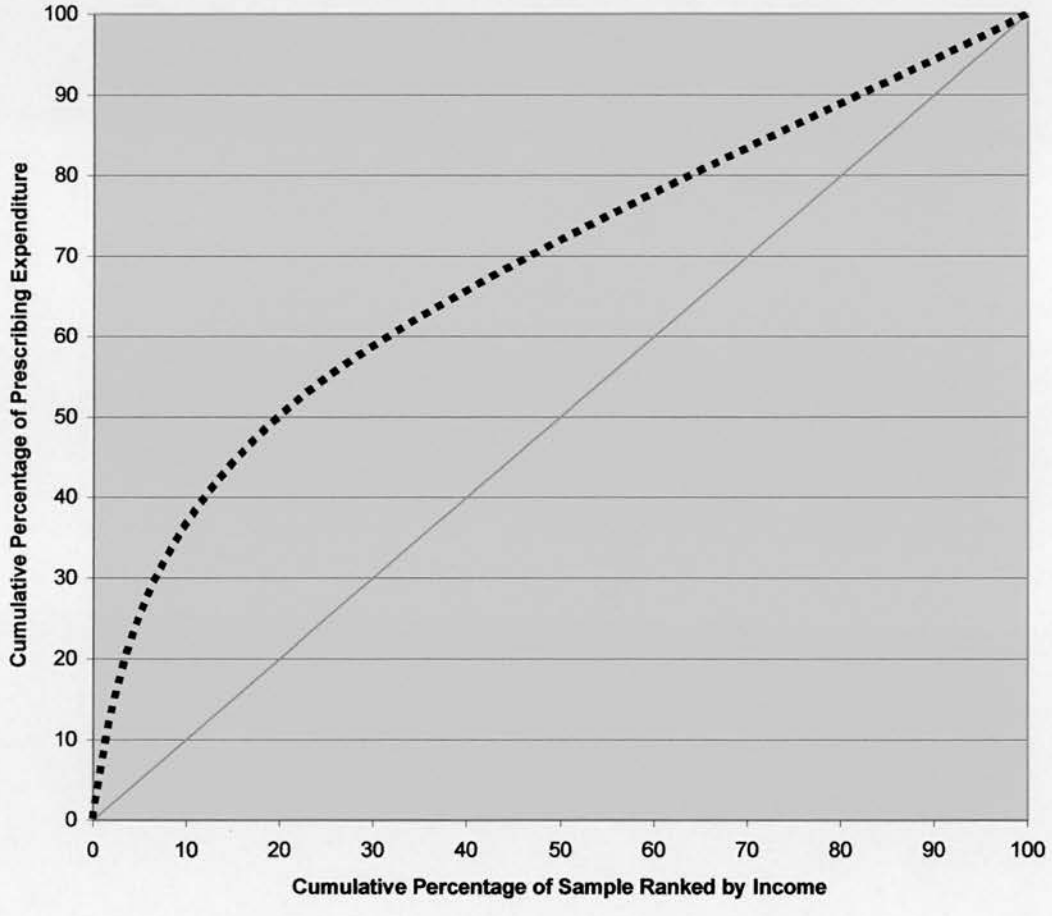
In addition, we consider the relationship between a model's predictions and income. A model that has a greater explained variance but underpredicts utilisation of poor people may be worse than one with a poorer explained variance but which accurately predicts utilisation of poor people. Essentially, we may wish to weight model residuals, penalising models with high positive absolute residuals for poor people (or other policy-relevant groups).

A rich theoretical and empirical literature on the measurement of inequality (and the related concepts of poverty and social welfare) has evolved over the last thirty years or so. This literature has focused mainly on income inequality, but can be applied to inequality in any vector, including health or health care utilisation. As well as univariate measures of inequality such as the Gini coefficient, a number of bivariate measures have been developed, which measure inequality in health or health care utilisation with respect to income or socioeconomic status. Of these bivariate measures, Wagstaff et al. (1991) recommend using the concentration index or relative index of inequality to describe income- or socioeconomic-related health inequalities. Most examinations of inequality in health care utilisation use the concentration index (Wagstaff et al., 1997; van Doorslaer et al., 2001). These are concerned with the measurement of horizontal equity rather than vertical equity, assuming that the implicit social welfare function is vertically equitable. In other

words, they assume that on average the health care system gets it right, and then deviations from the average for particular socio-economic groups are measured.

A concentration index describing income-related inequality in prescribing expenditure is illustrated in Figure 6.1.

Figure 6.1: Concentration Curve of Prescribing Expenditure v Income



Individuals are ranked in terms of income on the X-axis. If there is perfect equality of prescribing expenditure with respect to income, then the concentration curve would be on the 45^0 line, meaning that the bottom 10% of income earners incurred 10% of prescribing expenditure and so on. However, if the curve is above the line, then there is a pro-poor distribution of prescribing expenditure, meaning that the bottom 10% of income earners incur more than 10% of prescribing expenditure. This would be expected, since the presence of socio-economic health inequalities are well established. The concentration index measures the area between the concentration curve and the 45^0 line, with a value of zero indicating an even spread

of prescribing expenditure, positive values indicating a pro-rich distribution and the converse for negative values.

We borrow from van Doorslaer et al. (2001) and Kakwani et al. (1997) to compute concentration indices. The concentration index measures the area between the concentration curve and the 45^0 line of Figure 6.1, as follows:

$$C_y = 1 - 2\int_0^1 L_y(R)dR \qquad (6.3)$$

where C_y is the concentration index for prescribing, $L_y(R)$ is the concentration curve for prescribing as illustrated in Figure 6.1, that is the graph of the cumulative proportion of prescribing expenditure against the cumulative proportion R of the sample, ranked by income. A value of zero indicates that prescribing expenditure is spread evenly across the distribution of income, while a negative value indicates that the poor have a higher than average level of prescribing expenditure. A positive value indicates a pro-rich distribution of prescribing resources. The concentration index can be measured as:

$$C_y = \frac{2}{\mu_y}\mathrm{cov}(y, 1 - F(x)), \qquad (6.4)$$

where y is prescribing expenditure, $F(x)$ is the cumulative distribution of income, and μ_y the mean of prescribing expenditure and cov(.) is covariance.

6.4 RESULTS

6.4.1 Prediction Analysis

We wish to minimise prediction error. Prediction statistics are presented in Table 6.1.

Table 6.1

Comparative Measures of Prediction Error

Model	Prediction $\overline{R}^2$%	RMSE	MAPE	% Negative Predictions
NARA	12.99 (0.02)	315 (0)	191 (0)	
Demographic	14.41 (0.02)	313 (0)	193 (0)	8.02 (0.02)
Chronic Illness	38.77 (0.02)	268 (0)	154 (0)	6.76 (0.02)
Supply	38.93 (0.02)	268 (0)	154 (0)	6.63 (0.02)
Ln-Homo	34.61 (0.04)	336 (0)	179 (0)	
Ln-Het	35.42 (0.04)	279 (0)	157 (0)	
GLM	36.40 (0.04)	281 (0)	156 (0)	

Standard error of 50 random samples in parentheses

The best models in terms of prediction $\overline{R}^2$, in rank order, are the Supply model, the Chronic Illness model, the GLM, the Ln-Het, the Ln-Homo, the Demographic and the NARA. The variation about the mean $\overline{R}^2$ is very slight, such that all means are statistically different from each other. The prediction $\overline{R}^2$ varies from 12.99% to 38.93%, which is quite high for these type of models, but as outlined in Chapter 5, we expect that this is because age explains a lot of variance because since 1994 budgets have been set on the basis of age.

The Supply and Chronic Illness models have the lowest RMSEs and cannot be distinguished statistically, again illustrating their ability to minimise prediction error. These are followed by the GLM and Ln-Het. The NARA is second lowest, followed by the Ln-Homo, which has a very high RMSE. This is surprising, given its reasonably high $\overline{R}^2$. Absolute errors are minimised in Supply model, followed closely by the Chronic Illness, followed by the GLM and Ln-Het. The NARA is again second highest, followed by the Demographic model.

The proportion of negative predictions are zero for NARA and the two-part models (which is as expected given the construction of two-part models) and highest for the Demographic model.

In summary, the NARA is either second poorest or poorest against all the above criteria except proportion of negative predictions. Moving from the NARA to either the Chronic Illness model or the Supply model leads to a substantial reduction in prediction error. The Chronic Illness model and Supply model are virtually indistinguishable. Where there are statistical differences between the two, these are of very small economic importance. The GLM is the best of the two-part models, while the heteroscedastic retransformation is a better than the homoscedastic one for all these criteria.

We wish to maximise prediction variance, so that predicted expenditures reflect the variation in actual expenditures. The results are presented in Table 6.2.

Table 6.2

Comparative Measures of Prediction Variance

Model	Mean		Std Dev		Min		Max	
Actual	*184*	*(0)*	*338*	*(0)*	*0*		*1,719*	
NARA	185	(0)	124	(0)	25	(0)	348	(0)
Demographic	185	(0)	131	(0)	-165	(1)	521	(0)
Chronic Illness	185	(0)	207	(0)	-136	(0)	1,754	(2)
Supply	185	(0)	207	(0)	-136	(0)	1,755	(2)
Ln-Homo	241	(0)	370	(0)	4	(0)	12,408	(218)
Ln-Het	193	(0)	232	(0)	5	(0)	7,028	(194)
GLM	184	(0)	224	(0)	2	(0)	6,196	(160)

Standard error of 50 random samples in parentheses

Actual expenditure in the prediction sample had an average of IR£184 and a standard deviation of IR£338, while they ranged from zero to IR£1,719.35. For the one-part models, prediction mean was slightly greater than actual mean, while it was almost exactly equal to the actual mean for the GLM. The Ln-Het model overestimated predicted mean by IR£9 and the Ln-Homo model significantly overestimated it, illustrating the bias in the Duan smear. Amongst the one-part models, the Chronic Illness and Supply models have a standard deviation and range that matches the actual distribution the closest. In addition, we cannot distinguish between these two models. The two-part models have larger variation. While the GLM has a lower standard deviation than that of actual expenditures, it has a larger range. Meanwhile the Ln-Homo model overstates variation considerably, while its maximum predicted value of IR£12,408 represents a significant overstatement and one that would surely give policy-makers cause for concern. In short, the GLM appears to be the best of the above models with respect to prediction variance, so long as policy-makers are

unconcerned that its maximum prediction exceeds the truncation point for any individual's expenditure by over IR£4000. Meanwhile, all its predictions are non-negative unlike the other 'contenders', the Supply and Chronic Illness models.

Table 6.3 applies the Mincer-Zarnowitz test, that is, it tests the hypothesis that actual expenditure equals expected expenditure for each model by regressing predicted values on actual values and testing that the regression has an intercept of zero and a slope of one for each model.

Table 6.3

Mincer-Zarnowitz Test

Model	Intercept	% of times significant	Slope	% of times significant
NARA	0.617	0	0.997	0
Demographic	-0.608	8	1.001	8
Chronic Illness	-0.698	0	1.002	18
Supply	-0.714	0	1.002	20
Ln-Homo	63.462	100	0.493	100
Ln-Het	29.153	100	0.816	100
GLM	24.391	100	0.845	100

We find that all one-part models performed well, while neither of the two-part models do so. Except for the NARA, whose intercept and slope are insignificantly different from zero and one respectively, we can reject the hypothesis that actual expenditure is equal to expected expenditure for all individuals. However, the slope is very close to one for the other one-part models. Meanwhile, the two-part models overpredict low-expenditure individuals and underpredict high-expenditure individuals, with the two-part logarithmic model with the Duan smear performing particularly poorly. The results are more intuitively described in Figure 6.2.

Figure 6.2 Mincer-Zarnowitz Equations for Ideal Model and Two-Part Models

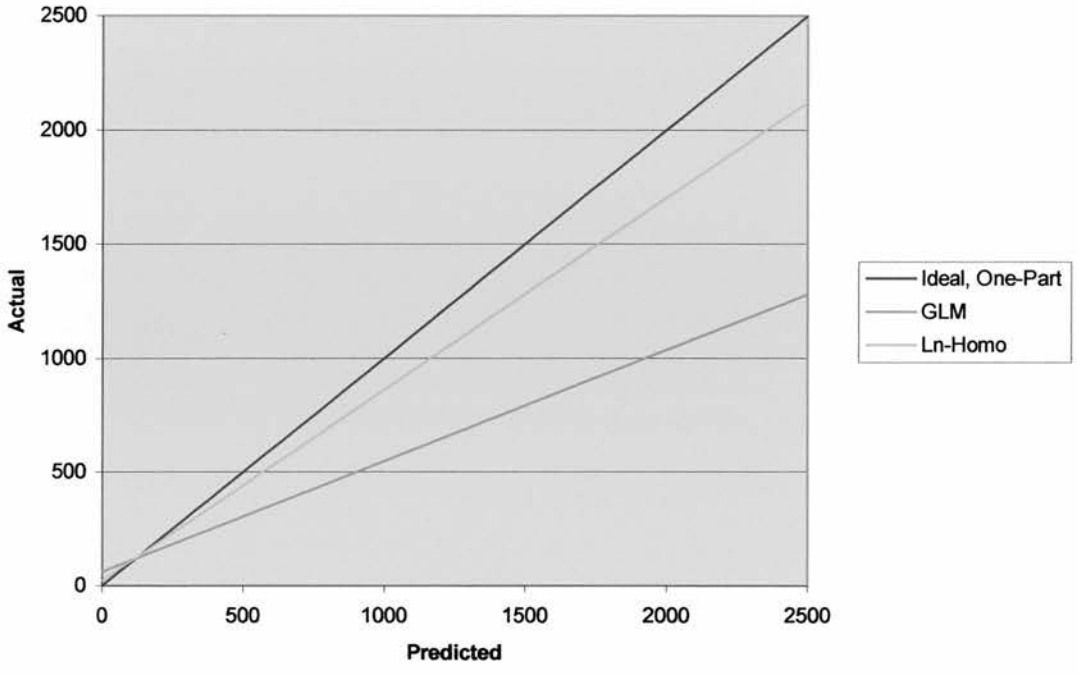


Figure 6.2 expresses predicted values on the X-axis and actual values on the Y-axis. The ideal model is one where the predicted values map exactly onto the actual values – when the predicted value is 500 the actual value is also 500. This is represented by the 45^0 line intersecting the Y-axis at the origin. While the one-part models are statistically significantly different from the ideal model (except the NARA), as described in Table 6.3, the two-part models overpredict low-expenditure users and underpredict high expenditure users. This may seem difficult to reconcile with the high predictions for certain individuals as described in Table 6.2. However, the median of the actual expenditures is IR£29.79, whereas the predicted value at this point for the two-part logarithmic model with homoscedastic smear is IR£72.99. Meanwhile, and partly due to it ability to generate negative predictions, the one-part model does not overpredict low-expenditure individuals to the same extent.

6.4.2 Distribution Analysis

This sub-section considers the distribution of prescribing resources with respect to income and vulnerable groups.

6.4.2.1 *Income-related Distribution of Need Predictions*

Table 6.4 describes the relationship between income and actual prescribing expenditure.

Table 6.4

Prescribing Expenditure by Income Quintile

Quintile	N	%	Income IR£	Prescribing Expenditure IR£
Q1	51,062	25.7	93	259
Q2	31,403	15.6	115	192
Q3	37,476	18.9	128	231
Q4	40,060	19.8	144	120
Q5	40,375	20.0	169	100

Five income quintiles were generated, with incomes varying from IR£93 to IR£169, that is, the poorest 20% of GMS individuals earned IR£93 per week, while the richest 20% earned IR£169 per week on average. As expected prescribing expenditure was inversely related to income, with those in the bottom quintile having an average prescribing expenditure of IR£259 as against IR£100 in the highest quintile.

We would prefer that prescribing models are more accurate for lower income groups as we do not wish these to be penalised by any model. Table 6.5 displays the prediction error by income quintile for each model.

Table 6.5

Comparative Prediction Error by Income Quintile

	Income Quintiles				
Actual - Predicted	Q1	Q2	Q3	Q4	Q5
NARA	15	-6	6	-1	-25
Demographic	4	-5	8	-4	-10
Chronic Illness	-7	-8	-2	-10	-13
Supply	-4	-4	2	-6	-9
Ln-Homo	-143	-65	-86	-10	-13
Ln-Het	-30	-15	0	4	3
GLM	-16	-2	9	11	7

We find that each model predicts actual expenditure by income quintile quite accurately. For the bottom quintile, the NARA underpredicts prescribing by the greatest amount (IR£15), while the Ln-Homo overpredicts for this group by IR£143. The Demographic, Chronic Illness and Supply models were most accurate for this group. The Ln-Homo model also overpredicts by the greatest amount for quintiles 2

and 3, which is largely due to its overprediction across most of the expenditure distribution as we can see from its considerable overprediction of the mean in Table 6.2.

The analysis by quintile shares in the table above does not produce an overall measure of income-related inequality in prescribing expenditure. A concentration index, however, provides such a measure. Concentration indices for each model are outlined in Table 6.6.

Table 6.6

Comparative Concentration Indices

Model	Concentration Index	
Actual Expenditure	-0.1678	(0.0003)
NARA	-0.1308	(0.0002)
Demographic	-0.1495	(0.0001)
Chronic Illness	-0.1578	(0.0002)
Supply	-0.1576	(0.0002)
Ln-Homo	-0.2090	(0.0002)
Ln-Het	-0.1983	(0.0002)
GLM	-0.1943	(0.0002)

Standard error of 50 random samples in parentheses

We find that the concentration index for actual expenditure was -0.1678, indicating a pro-poor distribution of prescribing resources – the concentration curve is above the 45^0 line, like in Figure 6.1 – which is consistent with the pattern described in Table 6.4. The current NARA model is considerably less pro-poor than the actual distribution of expenditure, while the Demographic, Chronic Illness and Supply models are slightly less pro-poor than actual expenditiure. The Supply and Chronic Illness models are statistically indistinguishable. The two-part models are more pro-poor than current expenditure. On the basis of Table 6.6, the Supply and Chronic Illness models are closest to the actual concentration index and are therefore the preferred options, unless we believe that the current actual expenditure is not pro-poor enough, in which case the two-part models are favoured.

6.4.2.2 Distribution of Predictions for Vulnerable Groups

Prediction error for each model by vulnerable group is split into Table 6.7, covering high expenditure groups, Table 6.8, covering low expenditure groups and Fig 6.3, covering chronic illness indicators.

Table 6.7

Comparative Prediction Error by High Expenditure Vulnerable Groups

	Disability	People Living Alone	Chronically Ill	Hardship Cases
NARA	84 (0)	-5 (1)	127 (1)	88 (1)
Demographic	-6 (0)	-8 (1)	112 (1)	87 (1)
Chronic Illness	-6 (0)	-14 (1)	65 (1)	51 (0)
Supply	-6 (0)	-15 (1)	64 (1)	51 (0)
Ln-Homo	-84 (0)	-162 (1)	-60 (1)	-35 (1)
Ln-Het	-28 (1)	-61 (1)	37 (1)	40 (1)
GLM	-7 (1)	-35 (1)	54 (1)	57 (1)
Actual	*275 (0)*	*305 (1)*	*360 (1)*	*310 (1)*

Standard error of 50 random samples in parentheses

Individuals from households in receipt of disability payments cost IR£275 on average. The NARA underpredicts disabled expenditures by IR£84, while the Ln-Homo and Ln-Het models overpredict it by IR£84 and IR£28 respectively. Other models were reasonably accurate, while the Demographic, Chronic Illness and Supply models are statistically indistinguishable.

Actual prescribing expenditure for people living alone is IR£305. Because they are an elderly group, the NARA is very accurate in predicting their prescribing need. The Demographic model is also very accurate, followed by the Chronic Illness and Supply models (which are statistically indistinguishable). The two-part models overpredict their expenditures, by a considerable margin in the case of the Ln-Homo model.

The average prescribing expenditure for people with chronic illnesses is IR£360. No model predicts their expenditures particularly well. The NARA and Demographic models underpredict expenditure considerably, while Chronic Illness and Supply models underpredict by approximately the same amount as the Ln-Homo model overpredicts it. The most accurate model is the Ln-Het, followed by the GLM, which underpredict by IR£37 and IR£54 respectively.

The average prescribing expenditure of hardship cases is IR£310. The Demographic model and the NARA underpredict their expenditures considerably, followed by the GLM, the Chronic Illness and the Supply model. The Ln-Het underpredicts their expenditures by approximately the same amount as the Ln-Homo overpredicts them. The most accurate model of hardship cases in the Ln-Homo, however.

Table 6.8

Comparative Prediction Error by Low Expenditure Vulnerable Groups

Model	Lone parent		Asylum Seekers		Unemployment Assistance		SWA		Early School Leavers	
NARA	-15	(0)	-69	(1)	-37	(0)	-12	(0)	-58	(1)
Demographic	2	(0)	-70	(1)	-26	(0)	0	(0)	-30	(1)
Chronic Illness	1	(0)	-48	(1)	-16	(0)	-1	(0)	-20	(1)
Supply	1	(0)	-48	(1)	-16	(0)	-2	(0)	-20	(1)
Ln-Homo	-8	(0)	-44	(1)	-20	(0)	-7	(0)	-7	(1)
Ln-Het	-5	(0)	-56	(1)	-29	(0)	-11	(0)	-23	(1)
GLM	7	(0)	-32	(1)	-5	(0)	12	(0)	-6	(1)
Actual	68	(0)	30	(1)	97	(0)	103	(0)	54	(1)

Standard error of 50 random samples in parentheses

Individuals from households in receipt of lone parent payments had an average prescribing expenditure of IR£68. All models are fairly accurate, with the NARA performing worst followed by the two-part models, and the Chronic Illness model and Supply model performing best.

Despite claims that asylum seekers are high users of health services, their prescribing expenditure is only IR£30 per person. All models overpredict their use, with the most accurate models in rank order being the GLM, the Ln-Homo, the Supply model and Chronic Illness model (jointly), the Ln-Het, and finally the NARA and Demographic model (jointly). Meanwhile, average prescribing expenditure for people in households in receipt of unemployment assistance is IR£97. The GLM minimises prediction error, followed by Chronic Illness and Supply models. The NARA has the largest prediction error.

Prescribing expenditures of people in households in receipt of supplementary welfare allowance is IR£103 and most models predict their expenditures well. The best model is the Demographic model while the worst are the NARA and the GLM. Finally, average prescribing expenditure for early school leavers still dependent on their parents was IR£54. Prediction error for this group was minimised by the GLM, followed by the Ln-Homo model. The NARA again had the highest prediction error.

Figure 6.3 illustrates the prediction error for each model for chronic illnesses, as measured by chronic disease scores.

Figure 6.3 Prediction Error for Chronic Illness

Actual average expenditures for cardio-vascular disease, epilepsy, rheumatological conditions, diabetes, glaucoma, respiratory disease, thyroid disease and psychiatric illness are IR£514, IR£562, IR£692, IR£737, IR£609, IR£596, IR£473 and IR£622 respectively. The general pattern for these variables is that the NARA and Demographic models underestimate actual expenditures by between IR£200 and just over IR£400; the Chronic Illness model and Supply model are extremely accurate and statistically indistinguishable; the Ln-Het and GLM are reasonably accurate, while the Ln-Homo model overpredicts by between IR£224 and IR£700. However, it is almost by construction that the first two models perform poorly, as these variables are included in the other four models but not in the first two. Nevertheless, we had no *a priori* expectation that the one-part models would outperform the two-part models, which they do spectacularly.

6.5 DISCUSSION

We make five points for discussion. First, in much of empirical economics, selection of models concentrates on relative explained variance or average prediction error. This chapter has applied a battery of performance tests to competing models to provide additional information to the model selection process. We consider not only average prediction error, but also prediction variation, prediction bias across the distribution of expenditure, which represents good practice in prediction analysis according to McCracken and West (2001). We also assess the distributional consequences of each model, with respect to income and policy-relevant variables. This level of rigour is novel in European risk-adjustment studies. In addition, we exploit the large dataset to apply split-sample estimation, which we repeat 50 times so that predictions have both point estimates and standard errors. Although this multiple sampling has appeared in a number of US risk-adjustment papers (Hornbrook and Goodman, 1995; Hornbrook et al., 2001; Deb and Burgess, 2002), it has not appeared in any European ones, as far as we are aware.

Second, the difference between the Supply model and Chronic Illness model is very small or non-existent for many criteria, suggesting that supply-side indicators do not affect need predictions to any great extent. If supply-side variables are difficult to collect, then risk-adjustment exercise could probably get away with excluding them without generating excessive prediction bias or inefficiency. Third, the standard

errors about prediction means were tiny, illustrating the stability of models estimated with such large sample sizes. Fourth, there is no evidence of unmet need. This is the first time that it has been tested for on individual-level data, as far as we are aware, so its absence is an interesting finding.

The remainder of the discussion section relates to model selection. Except for the Mincer-Zarnowitz test, NARA is second worst or worst against all tests of prediction error, the concentration index and prediction error for vulnerable groups. A switch away from the NARA is recommended.

The Demographic model is also unimpressive. Its prediction error is comparatively high, it is less pro-poor than actual expenditure or other models and it is dominated on many policy-relevant variables, especially the high expenditure ones. Thus the addition of demographic, socio-economic and access variables is not a significant improvement over the NARA. This is similar to many other risk-adjustment studies, as described in Chapter 2.

In addition, the Ln-Homo model performs poorly. Although it has reasonable prediction error properties, with a relatively high prediction R^2 and reasonably low MAPE, its RMSE is very high. It also overpredicts expenditure on average, and for high expenditure individuals, but especially for low-expenditure individuals as described in the Mincer-Zarnowitz test. In addition, it overpredicts for many policy-relevant variables including the chronic illness indicators. The presence of heteroscedasticity in the logarithmic scale is likely for much health expenditure data, leading to considerable bias if a homoscedastic retransformation is applied. It should be tested rigorously before application.

Given the poor performance of the above three models, we concentrate the discussion of model selection on the remaining models. The performance of the Chronic Illness model and Supply model is virtually indistinguishable, and indeed is statistically indistinguishable for RMSE, MAPE, standard deviation of the predictions, the concentration index and all vulnerable groups except those on Unemployment Assistance. Where they differ, the Supply model is the better performer.

These two models outperform the GLM and Ln-Het with respect to measures of prediction error, and especially prediction $\overline{R}^2$ and RMSE, as these penalise large errors more than small errors and the GLM and Ln-Het produce some very large errors, as exemplified in Table 6.2, where they produce maximum predicted values of over IR£6,000 and over IR£7,000 respectively.

While the Supply and Chronic Illness models generate a significant proportion of total predictions that are negative, the GLM and Ln-Het produce none, so they are likely to be favoured against this criterion.

With respect to prediction variance, the Ln-Het and GLM produce standard deviations first and second closest to the actual standard deviation, but they also produce a range in excess of the actual range. It is likely that policy makers would rate poorly two estimators that produce predictions of over IR£6,000 and over IR£7,000 for an individual when the effective ceiling of expenditure for the budget holder is IR£1719.35. Thus, despite the superior standard deviation of the Ln-Het and GLM, the Chronic Illness/Supply models may be favoured on this criterion.

The Mincer-Zarnowitz test also favours the Chronic Illness/Supply models. It exhibits how poor the GLM and Ln-Het are at predicting expenditure for low expenditure patients. However the Chronic Illness/Supply models are better able to predict for this group partly because they are generate negative predictions for some low expenditure patients.

The concentration index also favours the Chronic Illness/Supply models, unless we believe that the current concentration index for actual utilisation is not pro-poor enough. They also have the lowest for all high cost vulnerable groups except the Chronically Ill and hardship cases, which are better predicted by Ln-Het.

In conclusion, not even the NARA is dominated on all criteria that we have applied. Therefore, the choice of best estimator is a value judgement, depending on the relative weight that policy makers wish to place on each criterion. Nevertheless, given that the Chronic Illness/Supply models perform best against almost all criteria,

we favour either of these. Since amongst the criteria where there are differences in performance between these two models the Supply model performs best, we believe that it should be the preferred model.

The emergence of a one-part model as the preferred approach for risk-adjustment supports the views of van de Ven and Ellis (2000) and Diehr et al. (1999) as outlined in Chapter 2. However, we cannot claim that it will be the preferred approach in all instances and echo Mullahy (1998) who recommends that both one- and two-part models be estimated and tested empirically.

In unreported work, we performed the above tests on the full distribution of prescribing expenditure, with the same qualitative results emerging.

The performance of the Supply model compared with the NARA is worth summarising. The Supply model explains more than three times more variation in expenditure than the NARA; its average absolute prediction error is 20% lower than the NARA; it has greater variation in predictions, making it more responsive; it is more pro-poor; its prediction error for chronic illness indicators varies from less than 1% to 8% of that of the NARA; when we add the absolute prediction errors of all other vulnerable groups, we find that the errors of the Supply model are 45% of the errors of the NARA. Its only relative disadvantage is that it makes negative predictions for some individuals. However, at GP level these are likely to be compensated by positive predictions for others.

6.6 CONCLUSIONS

This chapter examined the need predictions of seven models, representing the six utilisation function estimated in Chapter 5, with two need models being generated from the logarithmic model, as two potential retransformations were suggested by previous literature. While many studies using utilisation functions, including risk-adjustment studies, apply explained variance, and average prediction error as performance tests, we also assessed prediction variance, prediction bias and the number of negative predictions. These represent novel approaches to model selection, at least in European risk-adjustment. In addition, we examined the distribution of need predictions by income, as measured by concentration indices and

quintile shares and the ability of each model to predict accurately for vulnerable groups.

Chronic Illness model and Supply model were the best against most criteria and were virtually indistinguishable. Where they were different, the Supply model was better, so we recommend it to model need for GMS prescribing. Meanwhile, the NARA was found to be either the poorest or the second poorest model against almost all criteria and a shift away from the NARA is strongly recommended.

Although no evidence of unmet need was found in the Appendix relating to this chapter, more rigorous tests could be designed. This is an area for future research.

7. CONSEQUENCES FOR GP PRESCRIBING BUDGETS

7.1 INTRODUCTION AND METHODS

In Chapters 5 and 6, we exploited individual-level data to estimate the determinants of prescribing expenditure and evaluate the predictive performance and distributive consequences of competing models. This is the same unit of analysis as used in the current IDTS budget setting scheme and is in line with most risk-adjustment exercises internationally, where such data exist. It allows us to calculate the effect of marginal panel changes on a GP's prescribing budget. However, budgets are set at the GP level rather than the individual level. Therefore, it is interesting to examine the effect of alternative budget-setting arrangements at this level of aggregation. The chapter covers five areas: prediction, distribution, risk exposure, effect of changing budget setting regime and case study analysis. As such this chapter tests a number of hypotheses with the unifying theme of using the GP panel as the unit of analysis. It can be read as a 'policy analysis' chapter.

This analysis includes all GPs whose GMS panels did not change by more than 10% in the period December 2000 to October 2001, which is the same inclusion criterion used in the Supply model. As such we did not calculate budgets for GPs who retired over the period or began accumulating other GPs GMS panels as these data distort the summary statistics.

The methods applied in each of the five areas are now covered in turn. First, this chapter begins by assessing the predictive performance and distributive effects of each estimator at the level of the budget holder. Next we consider risk exposure of the budget holder. Prescribing expenditure cannot be predicted with certainty, as the preceding chapters illustrate. The degree of risk exposure is dependent on the size of the budget and treatment of high expenditure patients. We examine the reduction in risk with increases in panel size and with the truncation of patient expenditures at different levels of truncation. Finally, we highlight the potential effect of this study on two practices, using qualitative and quantitative techniques.

We expect that prediction error as a percentage of the budget will fall with aggregation from individual to GP panel level, as negative and positive prediction

errors cancel each other out. We report the prediction R^2 to illustrate the effect of aggregation to GP panel level on explained variation out-of-sample. We also compare prediction variation across estimators at GP panel level similar to the individual level analysis in Table 6.2.

Individuals may aggregate into GP panels non-randomly. Therefore, while an estimator may pass the Mincer-Zarnowitz test, by being an unbiased predictor across the entire distribution at the individual-level, if individuals aggregate into biased groups, it may fail the Mincer-Zarnowitz test at the level of GP panel. This is tested empirically.

Second, the preceding chapter also examined the distributive consequences of each estimator. However, prescribing budgets are devolved to GPs, not individuals, so the distributive effects at GP level are arguably more important to measure. We describe the relationship between per capita income and per capita prescribing using quintile shares, as well as concentration indices for each estimator and for actual expenditure at GP panel level.

Third, one of the motivations for outlier removal discussed above was that one high expenditure patient could consume a large proportion of a GP's prescribing budget and so as to insure the GP against such risks, the budget for high expenditure patients should be set at a higher level of aggregation than the GP panel. In the GMS scheme, this higher level of aggregation is the GMS (Payments) Board, as budgets are not devolved first to a meso-level such as health boards and then to GP panels. In section 2.3.2 we examined a number of risk management techniques used in health care finance. We found that the IDTS uses two such techniques. First, there is the identification of certain drugs as budget neutral and their removal from GP budgets. Second, there is truncation of expenditure at a certain ceiling level, such that GPs only assume budgets for patients up to that level. In Chapter 4 we examined a number of levels of truncation of expenditure. In this chapter we examine the reduction in risk exposure of the GP with greater truncation as well as with and without the inclusion of budget neutral drugs in the GP's budget.

We wish to examine the effect of truncation while holding constant GP panel size. Therefore we generate a 'synthetic' GP practice of 800 individuals, randomly drawn from the prediction sample. This is repeated 200 times to account for variation in draws. In other words we bootstrap without replacement 200 times. In addition, we need to choose one model only. As described in the last chapter, the candidate models are the Chronic Illness, Supply and GLM. We have a slight preference for the Supply model, which we apply in this instance. The results of this exercise for either of the other two estimators are likely to be very similar to those for the Supply model.

Risk exposure is measured using standard deviation and percentiles of the bootstrapped distribution as a percentage of the GP's budget. They are examined under the following schemes: no truncation; truncation at the 99.9^{th} percentile; current IDTS scheme, which is the exclusion of budget neutral drugs and truncation at IR£2075; the exclusion of budget neutral drugs and truncation at IR£1719.35, which is the highest value of a 'non-outlier' according to Cooks D; the exclusion of budget neutral drugs and truncation at IR£1170.90, which is the highest level of expenditure getting a non-zero weight in iteratively-reweighted least squares and the exclusion of budget neutral drugs and truncation at IR£935.14, which is the 95^{th} percentile of expenditure.

An alternative risk management technique discussed in section 2.3.2 is the reduction in risk with increases in budget size. The generation of synthetic GP practices allows us to examine the relationship between risk exposure and panel size. Table 3.7 showed that while average panel size was 928, this varied from less than 50 patients to almost 1,800. The lower the number of patients in a GP's panel, the greater the expected variability in budgets, as the effect of one high expenditure patients cannot be moderated by a large number of lower expenditure patients. Indeed, it is unclear that the risk exposure of a GP with a panel of 1,800 is not excessive. Martin et al. (1997) illustrate the risks of setting acute hospital budgets for small areas of the UK. Therefore, we use the 'synthetic' GP panel idea to examine the increasing risks associated with reducing the budget holders population from 50,000 down to 100. The study of risk exposure represents some of this chapters more significant contributions.

Fourth, there is usually a cost to changing policy. If we found that the NARA and one of the other candidate models were the preferred options and exactly equal against every criterion, we would favour the implementation of the NARA, purely because it would not involve any costs in implementing a new formula. Hence, we examine the extent of change on GP budgets from the NARA to each of the competing models.

Finally, we conclude the empirical analysis with two case studies describing the effects of applying the ideas outlined in this study in two GP practices. We conducted face-to-face interviews with two GPs, both of whom have long running GMS contracts. The objectives of the interviews were to explore each GP's experience of the IDTS and their perceptions of health care needs in their own GMS panel. The first GP, John, comes from a materially deprived, inner-city area. We expect that his GMS panel are poor relative to the GMS average. The second, Joe, is based in a mostly middle-class suburban area. Although we do not know how affluent his GMS panel might be relative to the GMS average, we had a prior expectation that it was above average, based on his location. However, we acknowledge that this might not be true, as there may be a lot of hidden poverty in suburban areas. Both GPs were identified as having relatively progressive business practices. We supplemented these interviews with practice profiles based on the dataset generated for this study and examined the effect of applying each of the competing budget-setting models on their budget.

7.2 RESULTS

The results are reported in five sections, representing the five areas that the chapter examines as described above.

7.2.1 Prediction Analysis

There were 453 GPs in the prediction sample who did not experience a change in panel of more than 10% from December 2000 to October 2001. These had an average number of patients in the prediction sample of 371.

The prediction R^2 at GP level, by model, is described in Table 7.1.

Table 7.1

Comparative Prediction Error at GP level

Model	Prediction $\overline{R}^2$%	
NARA	77.79	(0.08)
Demographic	79.64	(0.08)
Chronic Illness	89.51	(0.06)
Supply	90.41	(0.04)
Ln-Homo	89.44	(0.04)
Ln-Het	89.39	(0.04)
GLM	89.78	(0.03)

Standard error of all bootstrap samples in parentheses

Aggregating from individual-level to GP panel level increases R^2 considerably. The NARA increased from approximately 13% to 77%, while the Supply model has the highest R^2 at over 90%. This is followed, not by the Chronic Illness model as was the pattern in the individual-level analysis, but by the GLM, then the Chronic Illness, then the Ln-Homo, then the Ln-Het. These final three are statistically indistinguishable.

Table 7.2 describes the variation in actual expenditure and predicted expenditure for each model, aggregated to GP panel level.

Table 7.2

Comparative Measures of Prediction Variance at GP level

Model	Mean	Standard Deviation	Min	Max
NARA	68,812	31,782	104	173,386
Demographic	68,775	32,036	170	175,362
Chronic Illness	69,634	35,019	170	184,537
Supply	69,982	34,313	170	190,765
Ln-Homo	94,295	48,998	125	271,631
Ln-Het	72,850	36,108	165	205,608
GLM	68,885	35,455	126	205,738
Actual	*68,764*	*37,766*	*3*	*208,003*

Aggregation to GP panel level produces a mean actual expenditure of IR£68,764, while the range about this mean is IR£3 - IR£208,003. The IR£3 represents a situation where only one patient from a GP's panel was in the prediction sample. As per the individual level analysis, predicted means are all greater than actual expenditure. The closest to the actual mean are, in rank order, the Demographic model, the NARA, the GLM, the Supply model, the Chronic Illness model, the Ln-Het and finally the Ln-Homo, which overestimates mean expenditure by a considerable degree. Similar to the individual-level result, the Ln-Het and GLM

produce standard deviations closest to that of actual expenditure. Unlike the individual-level result however, their maximum prediction is lower than the maximum actual expenditure, suggesting that the significant overprediction of the maximum value described in Table 6.2 at individual level does not translate into an overprediction of the maximum value at GP level. The one-part models all have lower standard deviations and lower ranges, while the poor performance of the Ln-Homo at individual level in Table 6.2 is also evident at GP-level.

Table 7.3 examines the effect of aggregation on prediction bias using the Mincer-Zarnowitz test.

Table 7.3

Mincer-Zarnowitz test at GP level

Model	Intercept	% of times significant	Slope	% of times significant
NARA	-1,407.68	0	1.08	100
Demographic	-1,639.17	0	1.08	100
Chronic Illness	-1,831.90	0	1.09	100
Supply	-1,078.40	0	1.07	100
Ln-Homo	3,498.69	100	0.74	100
Ln-Het	-745.08	0	1.04	60
GLM	-1,293.40	0	1.04	70

Standard error of all bootstrap samples in parentheses

All models have intercepts insignificantly different from zero except the Ln-Homo. All models have slopes that are significantly different from unity, for most samples. Except for the Ln-Homo, all models underpredict expenditure of low need GPs and overpredict expenditure of high need GPs. In addition, as with Table 7.2, the GLM and Ln-Het are much better at GP-level than at individual-level against this criterion. They have lower slope coefficients than the one-part models and smaller slopes.

7.2.2 Distribution Analysis

Table 7.4 describes the relationship between income quintile and prescribing expenditure at GP panel level.

Table 7.4

Prescribing Expenditure by Income Quintile at GP level

Quintile	N	%	Per capita Income IR£	Prescribing Expenditure IR£
Q1	90	19.8	154	210
Q2	89	19.7	160	196
Q3	91	20.1	163	185
Q4	90	19.9	166	177
Q5	93	20.5	171	157

Each income quintile has approximately 90 GP panels. The aggregation process by GP panel means that per capita income variation is much narrower than that exhibited in Table 6.4, where income quintiles ranged from IR£93 to IR£169. Equally, prescribing expenditure ranges from IR£210 to IR£157 in Table 7.4 as opposed to IR£259 to IR£100 in Table 6.4. Nevertheless, the negative relationship between per capita income and prescribing expenditure is clearly evident.

This relationship is measured using concentration indices in Table 7.5.

Table 7.5

Concentration Indices at GP level

Model	Concentration Index	
Actual Expenditure	-0.0523	(0.0004)
NARA	-0.0346	(0.0003)
Demographic	-0.0488	(0.0003)
Chronic Illness	-0.0635	(0.0003)
Supply	-0.0633	(0.0003)
Ln-Homo	-0.0800	(0.0006)
Ln-Het	-0.0709	(0.0003)
GLM	-0.0682	(0.0005)

Standard error of all bootstrap samples in parentheses

Actual expenditure has a pro-poor distribution as expected. It is not, however, as strong as that described in Table 6.7, because the aggregation process reduced the variation in income and prescribing expenditure. The NARA is less pro-poor than it 'should be', as is the Demographic model. However, the Demographic model is closest to the actual level of inequality. All others are more pro-poor than they 'should be'. In terms of proximity to actual inequality, they are ranked as Chronic Illness / Supply (jointly), GLM, Ln-Het and finally Ln-Homo.

7.2.3 GP Budget Risk Assessment

Table 7.6 exhibits the degree of risk associated with each level of truncation that was considered in Chapter 4 for a synthetic GP practice consisting of 800 randomly drawn patients repeated 200 times.

Table 7.6

Risk Exposure by level of Truncation of Expenditure

Truncation	Budget	Mean Residual as % of Budget	S.D. of Residual as % of Budget	2.5^{th} pctile	97.5^{th} pctile
None	168,147	-0.39	7.02	-14.56	14.44
99.9^{th} Pctile	166,732	-0.40	6.88	-14.50	13.76
IR£2075.00	149,438	-0.20	6.32	-12.84	12.37
IR£1719.35	146,284	-0.15	6.01	-12.50	12.25
IR£1170.90	136,717	-0.01	5.31	-10.10	9.45
IR£935.14	129,051	0.09	4.94	-10.20	9.21

When all expenditure is included, the average GP gets a budget of IR£168,147. At the 99.9^{th} percentile the budget drops by almost 1%. When the current IDTS scheme is applied, consisting of the exclusion of budget neutral drugs and truncation of expenditure at IR£2075, then budgets drop to IR£149,438, which is 89% of the budget with all expenditures included. Meanwhile, the budgets for the cut-off based on CooksD (IR£1719.35 and budget neutral drugs excluded), iteratively-reweighted least squares (IR£1170.90 with budget neutral drugs excluded) and the 95^{th} percentile with budget neutral drugs excluded (IR£935.14), respectively drop to 87%, 81% and 77% of the budget with all expenditures included.

Column three shows that the average expenditures in an 800 patient practice is roughly equal to average budget, with an average residual (expenditures less budget) of only 0.4% or less for all levels of truncation. Columns four, five and six then demonstrate the degree of variation in the difference between expenditures and budgets that are due to chance. For the 'no truncation' option, the standard deviation of the residual (the difference between actual expenditures and budgets) is seven percent of the budget. The interval between the 2.5^{th} percentile and 97.5^{th} percentile is (-14.56 to 14.44) meaning that 95% of deviations from budget should be less than 15%. The degree of variability in the residual falls with greater truncation, as expected. At the 95^{th} percentile, the standard deviation of the residual is just under five percent of the budget, while the middle 95 percentiles are no more than ±10% of the budget.

The second aspect of risk exposure that we consider, the relationship between panel size and variation, is described in Table 7.7.

Table 7.7

Variation in Budget by Panel Size

Panel Size	Budget	Mean Residual as % of Budget	S.D. of Residual as % of Budget	2.5th pctile	97.5th pctile
50,000	9,142,219	0.41	0.58	-0.71	1.53
25,000	4,570,440	0.37	0.83	-1.42	1.95
12,500	2,284,734	0.45	1.39	-1.96	2.89
7,500	1,370,235	0.43	1.80	-2.87	3.71
5,000	912,828	0.64	2.26	-3.42	5.24
2,500	456,386	0.73	3.15	-6.23	7.35
1,500	273,845	0.73	3.91	-7.63	8.13
800	145,978	0.80	5.41	-9.13	11.98
400	72,926	0.65	6.97	-13.54	13.93
200	36,196	0.50	10.66	-21.81	19.52
100	18,057	0.68	15.37	-26.93	27.29

Prescribing budgets vary from approximately IR£9.1m for a panel of 50,000 to approximately IR£18,000 for a panel of 100. While the mean residual is expected to be zero, this exercise found that residuals varied from 0.37% to 0.80%, which is reasonably close to the expected realisation. The measures of variation clearly demonstrate the increased risk borne by budget holders as panel size falls. For a panel of 50,000, standard deviation of the residual is less than 1% of budget, while the middle 95 percentiles vary by only –0.7% to 1.53%. However, for panels of 1500, 800 and 100, standard deviation rises to 3.91%, 5.41% and 15.37% of their respective budgets and the interval in the middle 95 percentiles increases considerably.

7.2.4 Extent of Change

Table 7.8 examines the effect of changing from the NARA to each of the competing models. GPs are ranked by change in budget and placed in percentiles, such that the data for the 1st percentile represents the ratio of the budget under each model to the budget under the NARA for the 1% of GPs whose budgets are reduced the most.

Table 7.8

Ratio of each budget to the NARA by budget setting regime

	Percentile					
Model	1%	5%	25%	75%	95%	99%
Demographic	0.91	0.94	0.97	1.00	1.03	1.07
Chronic Illness	0.53	0.79	0.98	1.05	1.11	1.20
Supply	0.50	0.78	0.96	1.03	1.09	1.19
Ln-Homo	0.36	0.88	1.24	1.38	1.50	1.71
Ln-Het	0.43	0.80	0.97	1.13	1.29	1.41
GLM	0.38	0.73	0.95	1.10	1.23	1.33

For most GPs the change in budget is far from dramatic. The inter-quartile range is only 3% for the Demographic model. The top 1% of GPs under this model only gain 7% and the bottom 1% of GPs lose only 9%. As is clear from other comparative tables in Chapters 6 and 7, the Demographic model is very similar to the NARA. The Chronic Illness model is a little more redistributive. The inter-quartile range is 7%, while the top 1% of GPs gain 20% and the bottom 1% of GPs drop to 53% of their NARA budget. However at the 5th percentile the drop is only to 79% of the NARA budget. The Supply model has a similar level of redistribution to the Chronic Illness model. The inter-quartile range is again 7%; the top 1% of GPs increase their budget by 19% and the bottom 1% of GPs suffer a drop to 50% of the NARA. At the 5th percentile the drop is only the 78% of NARA however.

The two-part models are more redistributive. The inter-quartile range for the Ln-Homo model is 14%; the top 1% increase their budget by 71%, while the bottom 1% of GPs drop to 36% of their NARA budget. This is the most redistributive of all models, but is biased by the fact the retransformation from the logarithmic scale produces an overestimate of the mean, implying that the predicted budgets for each GP exceed the total budget available for the full sample. This is illustrated by the result that both the 25th and 75th percentiles exceed unity.

The inter-quartile range for the Ln-Het model is 16%. The top 1% of gainers gain by 41% and the bottom 1% of losers drop to 43% of their NARA budget. At the 5th percentile, the drop is only to 80% while at the 95th percentile the increase is by 29%. Finally, the results for the GLM are also more redistributive than the one-part models. The inter-quartile range for the GLM is 15%; the top 1% of GPs increase their budget by 33%; the bottom 1% decrease their budget to 38% of the NARA and

at the 5th percentile the drop is to 73% of the NARA while at the 95th percentile the gain is to 123% of the NARA.

In unreported work we removed the six GPs who had GMS panels of less than 100 patients. This produced very similar results to those presented above. However, those GPs at the extremes of the distribution presented in Table 7.7 had smaller panel sizes than average.

7.2.5 Case Study Analysis

"The plural of anecdote is data"

Attributed to George Stigler

The interviews with John and Joe generated three broad themes. First, structural issues were discussed, such as the staffing of the practice and each one's perception of characteristics of his GMS panel. Second, we discussed business practices. Third, we discussed attitude to the Indicative Drug Target Scheme.

Case Study 1 - John

Interview

John has been a GP for 18 years, 16 of which he has had a GMS contract. He is in a single-handed practice. Staff consist of one half-time salaried GP, two secretaries, a part-time manager and a nurse attachment working on a clinical trail. A full time nurse is expected to be employed in the coming year. John believes that the practice's GMS panel is approximately 925, while the number of regularly attending private patients is about 660. As such, it has a higher proportion of GMS patients than the average GP, as expected for a GP based in a chiefly working class, urban area. John estimates that there are about 3-4 homeless patients, 8 asylum seekers and a small number of Traveller families. Other high need patients include chronic alcoholics, frail elderly and lone parents. John took over an older GP's list when he retired, so his perception is that it is a reasonably old list, even by GMS standards.

The practice has relatively modern business practices. It is based in a two-year old, purpose-built premises. Computerisation is used to manage prescribing, for patient summaries and sometimes in consultations. Between 2% and 10% of consultations are inputted into computer. Computerisation is set to further develop as a result of

recommendations in the Department of Health and Children's cardio-vascular strategy. Both John and the salaried GP attend Continuing Medical Education, and are members of the Irish College of General Practitioners, while the practice undertakes some clinical audit. While the practice does not advertise specialist clinics, each GP has particular specialisms.

John is in favour both of drug budgeting in principle and of the IDTS incentive scheme as well. His practice moved building and undertook a significant modernisation programme in 1999, which has allowed him to pay more attention to prescribing expenditure. Before the move he was in a badly managed, dilapidated practice that never made a drug budget saving. Since the move, he has made modest savings. He has misgivings about the loss of sight of the principle because the grant incentive is also included in the scheme. If GPs are not making savings on their budget, this discourages them from applying economy in their prescribing habits. Meanwhile he thinks that other GPs have found it possible to make huge savings on their budget and have received substantial practice improvements on the strength of them. This can reduce the morale of GPs who are attempting to make savings but are not able to do so because of the exceptionally high needs of their patients. For instance, he believes that there are GPs in deprived areas who are not able to ever meet their drug budget because of the poverty-related needs of their panels.

Practice Profile

Using the dataset constructed for this study, we can profile John's practice. Rather than report every variable, Table 7.9 reports those where John's panel differs significantly (t test, 5% significance) from the full sample average reported in Table 3.7.

Table 7.9

Characteristics of John's panel that Deviate from Sample Average

Variable	Practice Average	Sample Average
Age	40.33	43.45
Disabil	0.20	0.12
Lonepare	0.14	0.10
Marital	0.32	0.48
Numcard	2.23	2.43
Rural	0.03	0.61
Dist	1.04	1.68
Respir	0.07	0.05
Psych	0.09	0.07
Per capita income	154.02	163.24

John's practice is younger than the sample average. Indeed, only 22.8% of his panel is aged over 65 as against 28.3% of the full sample. The proportion of his panel from households in receipt of disability payments is strikingly different from the sample average as is the low proportion who come from household headed by a married person. The numcard variable suggests that smaller than average families attend his practice. Indeed, 48% of his panel have their own medical card, suggesting that he has above average number of people living alone. The distance variable is the average of the five distance indicators, where 1 is 'less than three miles from the GP' and 5 is 'more than 10 miles from the GP'. Ninety seven percent of John's panel live within 3 miles of the practice, illustrating its urban profile (which is confirmed by the rural residence variable). Thus, the panel can be characterised as relatively young, urban, with an above average number of single people, people living alone and small families. The panel's per capita income is in the bottom income quintile of the full sample. Despite the evidence of material deprivation, the only chronic illnesses that are above average are respiratory and psychiatric illness. This may be because the panel is relative young.

Implications of Altering Drug Budget Setting Formula

John's drug budget as a proportion of the NARA is described in Table 7.10.

Table 7.10

John's Drug Budget as a Proportion of NARA

Model	Proportion of NARA
Demographic	1.05
Chronic Illness	1.12
Supply	1.13
Ln-Homo	1.57
Ln-Het	1.32
GLM	1.24

John would gain considerably from the introduction of any of the above budget setting regimes. A comparison with Table 7.8 shows that John would be in the top 5% of gainers from a regime change.

Case Study 2 - Joe

Interview

Joe has 25 years experience as a GP, 20 of which he has had a GMS contract. The practice consists of two partners, an assistant GP, 2.5 whole time equivalent secretaries, and a research nurse attachment. The GMS panel is 771 and the number of private patients is not known. The GMS panel comes from a mixed urban – rural area. The panel is dispersed throughout the community, rather than being drawn principally from one local authority housing estate, for instance. The practice has an interest in psychiatric morbidity, and it perceives that it has an above average number of patients suffering mental illness as a result. No other groups of high need patients were identified.

The practice applies relatively modern business practices. Computerisation is used for administrative purposes, but not for prescribing or patient histories. Joe attends Continuing Medical Education and is a member of the Irish College of General Practitioners, while the practice undertakes clinical audits in suicide, diabetes and vaccinations, as well as participating in research projects. While there are not formal specialist clinics, the staff specialise in particular areas and promote these specialisms among their patients.

Joe considered that the original drug budgets were generous, so that generating savings was not difficult. The practice made savings on its budget previously which were used for practice renovation. The savings were accomplished by designing and

implementing a practice formulary. The formulary has now gone out of date and the practice budget is now in deficit. Continuing to keep the practice formulary up-to-date and paying attention to generic prescribing rates required co-ordination between all physicians in the practice, which they have not been able to achieve recently because of other priorities.

Practice Profile

Table 7.10 outlines those characteristics of Joe's practice that differ significantly (t-test, 5% significance) from the sample averages, as described in Table 3.7.

Table 7.11

Characteristics of Joe's panel that Deviate from Sample Average

Variable	Practice Average	Sample Average
Gender	0.63	0.55
Marital	0.37	0.48
Numcard	1.96	2.43
Rural	0.33	0.61
Dist	1.26	1.68
Glau	0.02	0.01
Psych	0.09	0.07
Per capita income	154.23	163.24
Specific[1]	0.53	0.48

NOTE 1: Not tested for statistical significance as only one observation

Joe has an above average proportion of his panel that are female and unmarried. A greater than average number of his patients have their own medical cards, suggesting that they come from small families or live on their own. As was his perception, the panel is mixed urban/rural, although predominantly urban, with 83% living less than 3 miles from the GP. As he suggested, he has above average number of patients with psychiatric illness. He also has an above average number with glaucoma. As with John's panel, per capita income in Joe's panel is from the lowest quintile of the sample average. The practice can be characterised as urban, female, unmarried and poor. The proportion of Joe's prescribing that is classified as specific is above average, putting him in the top 20% of GPs in terms of specific prescribing, which suggests that he is a 'good prescriber'.

Implications of Altering Drug Budget Setting Formula

Joe's drug budget as a proportion of the NARA is described in Table 7.12.

Table 7.12

Joe's Drug Budget as a Proportion of NARA

Model	Proportion of NARA
Demographic	1.04
Chronic Illness	1.07
Supply	1.08
Ln-Homo	1.41
Ln-Het	1.24
GLM	1.17

Although Joe's gains are not as great as John's, he would still experience reasonable increases in budget from a change of budget setting regime. The changes reported above put him above the 75th percentile in the distribution of gainers and losers reported in Table 7.8 for all models, except the Demographic model, where he is above the 95th percentile.

7.3 DISCUSSION

The chapter is concerned with a number of questions relating to the application of weighted capitation techniques to budget setting and can be considered a 'policy analysis' chapter. A number of questions were addressed: What is the effect of aggregation on predictive performance? What is its effect on distribution of prescribing resources? What are the effects of truncation and panel size on risk exposure? What is the effect of changing from the NARA on the redistribution of GP budgets? Finally, do the ideas discussed in this study make sense when applied to two GP practices? Six points for discussion emerge.

First, Table 7.1 finds that aggregation to GP panel level increased prediction R^2 to approximately 77% to 88%. This compares with an R^2 of approximately 95% for GP-level risk-adjustment in England (Rice et al., 2000). Meanwhile, Newhouse et al. (1989) were able to explain 60% of health care expenditures for groups of patients in the US. Rice et al. (2000) may have a higher R^2 partly because English GP practices are much larger than Irish ones. Single-handed GPs in the UK have panels of between 1,500 and 2,000 patients, while there are far more multi-partner GP practices. In addition, Rice et al.'s (2000) R^2 is an in-sample result, which is more susceptible to overfitting than the out-of-sample R^2 that we report. However, Rice et al. (2000) considered all prescribing expenditure, which should reduce explained variance, whereas we have truncated prescribing expenditure.

Second, aggregation improves the performance statistics of the two-part models, and especially the Ln-Het and GLM models. The prediction R^2 for the GLM is greater than that of the Chronic Illness model, while the other two two-part models are indistinguishable from the Chronic Illness model with respect to this statistic. Meanwhile, the high maximum predictions for the two-part models at individual level do not translate into excessive maximum predictions at panel level. Furthermore, the Mincer-Zarnowitz test for the GLM and Ln-Het indicate that they are slightly less biased than one-part models. They still underpredict at low levels of expenditure and overpredict at high levels, but to a lesser extent than one-part models. However, they deviate from the actual level of inequality by a greater degree than the Chronic Illness, Demographic and Supply models. While this study uses the individual as the unit of analysis and therefore model selection analysis should also use this unit of analysis, meaning that the Supply model is the preferred model, and while the Supply model is probably still the best performing model at panel level (with the highest prediction R^2 and the second lowest deviation from the actual concentration index), panel level analysis suggests that the two-part models behave reasonably well. They should continue to be considered as alternatives to one-part models in risk-adjustment studies.

Third, income inequality at GP panel level is less than that at individual-level as expected. It is still present, however, and the negative relationship between prescribing expenditure and income is still evident, although the concentration index at panel level is less pro-poor than that at individual level. However, we do not know if GPs are equally progressive across the sample. For instance, we do not know that the devolution of budgets to GPs will produce the concentration indices that we found in Table 6.6. It is possible that the GPs who gain are less progressive than those who lose, such that the ultimate realisation is a less progressive distribution of prescribing resources. Sutton and Lock (2000) found that health boards in Scotland differed in degree of progressivity and suggested that budget setting formula should account for inter-health board variations in progressivity. In Scotland, health boards receive prescribing budgets based on a GP-level risk-adjustment formula. Health Boards then have discretion to allocate budgets to GPs as they see fit. The Irish health service employs a similar resource allocation

mechanism at a lower level of aggregation – the Scottish health board is the Irish GP and the Scottish GP is the Irish patient. Therefore, we could measure the progressivity of individual GPs and use this information to monitor their distribution of prescribing resources, to ensure that they are sufficiently pro-poor. It is unclear, however, what the benchmark level of progressivity should be. We leave this area to future research.

Fourth, we use synthetic GP panels to assess the effect of truncation and panel size on risk exposure. A number of methodological points need to be noted. First, we chose standard deviation and percentiles of the bootstrapped distribution to describe risk exposure of the GP. Further work in this area should consider value-at-risk or the excess loss probability estimator to highlight risk exposure (Cotter, 2001). Second, the analysis is predicated on the assumption that patients randomly pick their GP. This may not be the case. GPs working in low-income areas may be more likely to get a high expenditure patient than a GP working in a high-income area. Therefore, for certain types of GPs the risk exposure may be greater than that expressed here and for others it may be less. There is no well developed theory of how patients are choose their GP, although there is evidence of very little 'shopping around' (Phelps, 2000), so the assumption of random selection appears to be a reasonable approximation. This assumption can be tested empirically. GP practices could be ranked by income, the number of high expenditure patients in each practice could be calculated and that vector could be tested for random walk. We leave it to future research to determine the suitability of the assumption of random assignment of patients to GPs.

Table 7.6 sharpens the focus on the trade-off between risk-exposure of the GP and efficiency of the scheme. With truncation at the 95th percentile and budget neutral drugs removed, only 77% of total GMS expenditure is covered by the scheme, but the average GP's budget is within 10% of their expenditures, 19 times out of 20. With the current scheme, 89% of total expenditure is covered by the scheme, while the average GP's budget is within 13% of their expenditures, 19 times out of 20. This intervals widens to 15% when all expenditures are included. Our choice in Chapters 4, 5 and 6 was to truncate at IR£1719.35 and exclude budget neutral drugs, meaning that the 'contaminating' effect of outliers is tempered to a large degree

producing a more efficient empirical model. Meanwhile, 87% of all expenditure is covered by the scheme and risk to the GP is reduced, as the GP's budget is likely to be within 12.5% of their expenditure 95% of the time, as against approximately 15% if all expenditures are included in the scheme. In the absence of information on risk preferences, we cannot make a positive statement on the optimal point on the risk/efficiency frontier. We leave an assessment of risk preferences to future research.

A clear relationship between panel size and risk borne by the budget holder emerges in Table 7.7. For panel sizes of only 100, a GP has a one in forty chance of having a panel whose expenditures exceed the GPs budget by 27%. Although the same upside risk exists, most people would generally wish to insure against excessive risks. If policy-makers and GPs are concerned about this level of risk-exposure, there are a number of alternatives. They could consider removing GPs with small panels from the GMS scheme. Removing GPs with less than 100 patients would remove only 0.5% of total expenditure from the budgeting scheme. Alternatively, GPs with small budgets could be given two- or three-year rolling budgets. Finally, budgets for a number of GPs could be amalgamated, thereby pooling risks. If these pooling structures were set up, then they could be used not only for GPs with small panels, but for all GPs to pool their high-expenditure patients. Instead of removing 11% of total expenditure from the scheme, as we have done by removing high-expenditure patients and budget-neutral drugs, meaning that the risks associated with these expenditures are borne nationally, budgets for high expenditure patients could be devolved to a meso-level, such as health boards, community care areas, or groups of GPs.

We consider only three risk-management techniques here, while six are reviewed in Chapter 2. We consider the removal of patients from the budgeting scheme once they reach a ceiling (also known as stop-loss arrangements), the assignment of certain drugs as budget neutral and the removal or pooling of budgets for GPs with small panels. Three other risk-management techniques that could be considered are the removal of high risk patients (which is the prospective removal of patients who are likely to be high expenditure, such as AIDS patients or the terminally ill), partial capitation and a combination of prospective and retrospective reimbursement. Future

research should consider the relative effectiveness of each of these risk management strategies for indicative drugs targets.

Fifth, we found that the introduction of alternative budget setting regimes do not alter budgets for many GPs dramatically. The top performing model – the Supply model – produces changes in budgets of –22% to +9% for the middle 90% of GPs. Although this does not seem particularly large, many GPs expenditure is within 10% of the NARA. For instance, category A prescribers (those GPs who exceed their budget by the greatest amount) were defined in 1998 as GPs whose actual expenditures for 1997 were in excess of 105% of the NARA; category B GPs had actual expenditures for 1997 of between 95% and 105% of the NARA and category C GPs had actual expenditures for 1997 of less than 95% of the NARA. Therefore an increase of 11% in budget could move a GP from the A category to the C category, so the redistribution described here is economically important.

Sixth, the case studies offer an interesting perspective on the policy implications of this study. John has an approximately average GMS panel size. It is interesting to note that he perceived that his GMS panel was relatively old, whereas in fact it is well below the GMS average. In addition, he did not possess data on the number of high need patients on his panel. For instance, the proportion of his panel from households in receipt of disability payments is well above average, while there are also above average levels of respiratory and psychiatric morbidity. He expressed no awareness of these. This illustrates the potential usefulness of informing GPs of their practice profile using the database constructed for this study.

John employs relatively modern business practices, including participation in research projects, clinical audit and Continuing Medical Education. He also recently moved to a new premises. Moreover, he expressed an interest in drug budgeting. If his budget reflects the relative needs of their patients, then he appears to be a likely candidate to make savings. However, John was unable to make a saving in his previous premises, while he has only made modest savings since the move. We find, however, that John's panel is a good example of the type of GMS panel whose expenditures are underestimated by the NARA. We found above that the NARA is not as pro-poor as alternatives, while John's practice is among the poorest in the

sample. Furthermore, the NARA does not reflect the disproportionate levels of disability and respiratory and psychiatric morbidity in John's panel. As a consequence, his budget is underestimated by 13% if we apply the Supply model. He expressed a concern that deprivation was not factored into the budget setting process appropriately, and his panel is a good example of the effect of that omission.

Finally, it is interesting to note the non-pecuniary effects of drug budget schemes. John believed that drug budgeting was 'the right thing to do' irrespective of financial rewards, while he stated that being over-budget had a negative effect on a GP's morale. It would be very unfortunate if the current scheme was damaging the morale of GPs serving deprived populations, as this group are crucial to achieving reductions in health inequalities.

Joe also has an approximately average GMS panel size. His perception of the particular needs of this patients was a little better than John's as he indicated that his panel had a disproportionate number of people with psychiatric illness. However, he expressed no awareness of the high prevalence rate of glaucoma in his panel, nor did he seem to know that his panel was particularly poor. Indeed, we picked Joe's practice because it is based in a predominantly middle-class suburb and we suspected that the GMS panel would be relatively affluent. However predicted income for the panel indicates that it is amongst the poorest in the sample. The outstanding finding of Nolan et al.'s (1998) study of the spatial distribution of poverty in Ireland was that it is widely dispersed, rather than concentrated in stereotypical 'sink' housing estates. Joe's GMS panel seems to be a good example of the same phenomenon, with very poor people living in an area that is generally perceived to be middle-class.

Joe also employs modern business practices, including participation in research projects, clinical audit and Continuing Medical Education. He also appears to be a 'good prescriber', as the proportion of his prescribing that is 'specific' is well above the sample average, while the practice had a prescribing formulary in the past. Indeed, he states that he found it easy to make savings on his budget at the beginning. This is despite the fact that the NARA underestimates the needs of this practice by 8%, if the Supply model is taken as the 'true' measure of need. There are two interesting points to note here. First, Joe's experience suggests that it is possible

to make savings on one's prescribing budget even when that budget is an underestimate, through the application of good prescribing habits. Second, the incentives in the scheme may not be strong enough to motivate GPs to make continual savings. Once Joe made savings and renovated his practice on the strength of them, the incentive to make further savings diminished. It would be interesting to know if this is a widespread phenomenon, and if so could the incentive structure be changed, for instance, by applying a sanction to GPs whose budgets are in deficit. Finally, one of the reasons that the NARA underestimates Joe's prescribing need is that Joe has a disproportionate number of people with psychiatric illness on his panel. He suggests that this is so because the practice is known for its expertise in psychiatric medicine. The NARA only rewards GPs who specialise in illnesses that are strongly to age, which excludes psychiatric illness. This is manifestly unfair and may generate unwanted risk selection strategies. Therefore a move to any of the three favoured budget setting models would counter this problem, reducing disease-related inequity and potentially risk selection.

7.4 CONCLUSIONS

This chapter examined the effect of risk-adjustment at the level of the GP panel. The results can be summarised into five points. First, aggregation allows high positive and negative residuals to cancel each other out so explained variance increases and prediction error is reduced. Second, aggregation leads to a modest amount of prediction bias. Third, it reduces the distributive effect of risk-adjustment. Fourth, increased truncation and increased panel size reduce the risk exposure of the GP. Fifth, changing from the NARA to alternative budget setting regimes does not have a large effect for the majority of GPs but this effect may be economically important. Sixth, the results of the empirical work in this study are sensible and useful when applied to particular GPs and can enhance their knowledge of their GMS panel.

Five areas for future research were also identified. First, we identified the need for an assessment of the relative progressivity of Irish GPs in distributing prescribing resources and the effect this may have on progressivity of overall budget setting exercises. Second, we would like to know if there is a systematic component in the pattern by which patients choose their GP, or can we assume random selection? Third, what are the risk preferences of GPs and the GMS (Payments) Board and what

are the implications of these for truncation and minimum budget size? Fourth, we suggested that new budgetary structures could be implemented, at a level between the GP and the GMS (Payments) Board, which could assist in risk management. The effect of such structures would be an interesting topic to research. Fifth, there are a number of alternative risk management strategies used elsewhere, which could have application to the IDTS. Their usefulness needs to be assessed further. Sixth, the role of financial and non-financial incentives in the indicative drugs target scheme could be assessed. Would GPs continue to adhere to prescribing budgets even in the absence of financial rewards? Would they adhere to them more closely if they were sanctioned for overspending? What are the effects of budgeting on morale in General Practice in Ireland? We are reluctant to make normative statements such as that these areas should be researched, especially when we do not know the opportunity cost of such research, so we leave it to other to decide their relative value.

8. CONCLUSIONS

This study examined the determinants of prescribing expenditure by recipients of General Medical Services in Ireland. A number of contributions were made.

In the Introduction we reviewed the normative approach to risk-adjustment and decided against using it in an Irish setting. In Chapter 2, we identified gaps in previous risk-adjustment literature and attempted to fill them. We proposed new measures of physician prescribing style. We reviewed approaches to managing risk exposure only some of which are currently used in the IDTS. In Chapter 3, we described the construction of the dataset used in the study. This included the 'discovery' of the Medical Card Register, the generation of chronic disease scores and a method for assessing their level of measurement error, and the generation of prescribing style indicators. We applied multiple imputation to missing records and imputed income for everyone in the study. In Chapter 4 we contributed to the debate on the specification of age in micro-econometrics, applied rigorous outlier identification techniques for the first time in risk-adjustment, examined the dataset using quantile regression and examined the usefulness of finite mixture models for risk-adjustment. Chapter 5 examined the determinants of prescribing expenditure, finding that chronic illness, age and disability were its primary determinants. We also contributed to a debate in health econometrics on the relative value of one- and two-part models and the appropriate way of specifying the second part of the two-part model. Chapter 6 generated capitations from the utilisation functions described in Chapter 5, and applied a number of tests of the prediction performance and distributive implications of each one, representing best practice in risk-adjustment. Chapter 7 examined predictive performance and distributive implications at the level of the budget holder. It also assessed the risk exposure of the budget holder, and described the effect on budgets of applying different risk-adjustment formulae to the IDTS. The implications of the study for two GPs were also described in two case studies.

In the Introduction we briefly reviewed the IDTS budget setting methodology. We found it lacked transparency. We suggest that a clearer budget setting methodology be applied. We also reviewed the 'normative' approach to risk-adjustment. We

found that it was not as popular as the 'empirical' approach, principally because it did not allow for a direct link between need and health care use. Moreover, data on SMRs and DALYs are not readily available for Ireland at a level of disaggregation that could be used for risk-adjustment in primary care. However, we suggested that it had a number of advantages, not least that it would not lead to a bias against socially excluded groups, and suggested that more research into the assignment of resource weights to levels of need was worthy of further research.

Following the critique of the standard risk-adjustment model by Schokkaert and van de Voorde (2000) and UK risk-adjustment studies in the tradition of Carr-Hill et al. (1994), two gaps in the risk-adjustment literature were identified in Chapter 2. First, a review of the expected effect of each covariate from first principles had not been undertaken. Thus, the basis on which variables were assigned 'counter-intuitive' status for budget setting was unclear. Second, since the standard risk-adjustment model does not include non-needs variables, a review of other studies applying health care utilisation functions was required, in order to determine what variables should be included in such a function and what previous literature can tell us about their expected effects. These were reviewed in Chapter 2 and expected effects were described. Therefore, this study is more transparent and comprehensive in how expected effects are generated than many other risk-adjustment studies.

Chapter 2 also found that attempts in previous studies to model physician practice style had met with limited success. In this study we wished to include a measure of the concept of relative aversion or partiality to prescribing by each GP. Therefore, we proposed new measures of physician prescribing style based on McGavock (1988). We constructed these indicators for the study population and applied them, finding that they were significant in most instances. The sign on the 'specific' variable was contrary to expectations in some cases, while the other two indicators were signed with expectations. They represent a promising development in the field and we recommend the use of these variables, or other variables representing relative prescribing aversion generated from McGavock's (1988) classification scheme, in future research. Further tests of their validity as measures of the underlying concept are also warranted.

A unique dataset was generated in order to model prescribing expenditure, as described in Chapter 3. This is the first time that Medical Card Register has been used for research purposes, as far as we are aware. The Medical Card Register was tested for errors and validated against external data. While we were only able to test for logical inconsistencies and to do validation tests at aggregate level, and while further testing of reliability of data capture and coding would be welcome, indications are that these data are of good quality. This is not surprising since the data in the Medical Card Register are generated from the GMS eligibility testing process, which is taken very seriously as it confers significant benefits on the recipient.

The Medical Card Register has proved in this study to be a powerful source of demographic and socio-economic data, as well as data on access to services, for the poorest and sickest groups in Ireland. As well as age, we found disability, marital status and access to services to be particularly important determinants of prescribing expenditure and lone parenthood surprisingly unimportant. The vulnerable groups data assisted us considerably in model selection. The availability of these data for the full sample would be a welcome improvement in the Medical Card Register. In addition to its use in this study, and as a consequence of this study, the Medical Card Register is now being used in other health utilisation research in Ireland including an examination of dentist utilisation by Woods (2002).

We found chronic illness indicators to be powerful predictors of prescribing expenditure, a finding that is mirrored in previous research on health care utilisation (Hornbrook et al., 2001; Lamers and van Vliet, 2001). As well as their use in this study, future research into the relationship between chronic disease scores and other aspects of health care utilisation in Ireland, such as GP consultations would be an interesting area of research.

In addition, although we found that the chronic illness indicators had lower estimates of prevalence rates of particular chronic illnesses than epidemiological estimates, further work on their use as measures of community morbidity is merited. As an alternative to epidemiological data, of which there is a paucity in Ireland, they are a

cheap, locally sensitive and timely. What remains to be determined is if they are also valid.

Concerns have been expressed with using diagnosis-based or prescribing-based measures of health status for risk-adjustment (Ellis, 2002). We proposed and applied a test of measurement error in chronic disease score data, testing the effect of GP characteristics, including GP prescribing style, on the probability of being indicated as a sufferer of a chronic illness. We found that the effect of the GP was tiny, supporting the use of these indicators for risk-adjustment.

Another unique element of the dataset is the set of variables on GP characteristics. We found that GP age and prescribing style were especially important determinants of prescribing expenditure. As a consequence of this study, these variables have been used by others in health economic research. For instance, O'Laughlin (2002) used the set of GP variables generated in this study to assess the effect of relaxing the budget constraint on statins on GP prescribing of statins.

Another contribution is the multiple imputation of missing data, which improved the precision of our results. Brownstone and Valletta (2001) report only one use of multiple imputation in empirical economics, while its use is growing steadily in many others (King et al. 2001), so this study offers a contribution to empirical economics.

Using the Household Budget Survey 1999-2000 (2002), we imputed income for every individual on the Medical Card Register. We found a significant difference in income between the 74% of all GMS recipient households where all members were entitled to GMS services and the remainder, where only a fraction of the household had GMS entitlement. The percentage of the household aged under 14 was associated with receipt of full GMS entitlement, as were households headed by a married person and households in receipt of disability payments. Thus these households are more likely to live in poverty, according to our estimates. Unlike previous research on poverty in Ireland which identified lone parents as being at a high risk of poverty, we find that households headed by a lone parent has no effect on income. This may be due to changes in economic conditions since the poverty

research cited above, which is worthy of further examination. The indicators of poverty identified above are also worthy of further comparison with previous poverty research in Ireland.

Previous research in empirical microeconomics did not provide any consensus on the specification of age for this study. We assessed a number of candidate specifications and found them to be statistically equivalent, which is an interesting methodological contribution.

Many previous risk-adjustment studies have removed outliers, but in an unsystematic way. We applied more rigorous outlier identification techniques, which took into account leverage as well as residual. We found a high proportion of the sample were identified as outliers, even using conservative outlier identification criteria, and that most of these had high expenditures. We chose to truncate the distribution, at a point informed by outlier analysis. However, it would be interesting if the reasons why certain observations are outliers could be identified. Careful examination of drugs included in each outlier's total expenditure may provide some answers. Given the sample size we work with, this was not possible for this study, but may be of worthy of consideration in smaller samples in future research.

The value of quantile regression for describing the conditional distribution for prescribing expenditure was also highlighted. We found that those at the 95^{th} percentile had a very different set of slope coefficients, highlighting that high cost patients were somehow different from the rest of the sample. The exercise also highlighted the degree of heterogeneity in chronic illness indicators, age and its square and disability especially. Further exploration of these variables may generate more refined measures of health care need.

We used the income data to examine the distribution of need predictions by each model. We found that the actual distribution of prescribing resources was pro-poor and found that the Supply model was the closest match to the actual distribution. Whether the current distribution is pro-poor enough is, however, an open question.

Since Duan (1983) there has been regular debate about the use of two-part models and the retransformation of the second part in health care utilisation data. We contribute to the debate in two ways. First, Manning and Mullahy (2001) outline a number of potential specifications for the second part of the two-part model. We apply the most popular one, the homoscedastic smear, as well as the two most promising alternatives. Second, Deb and Trivedi (2002) suggest that the finite mixture model provides a more flexible and responsive specification of health utilisation data than does the two-part model. We apply finite mixture models to risk-adjustment data. Both these applications are amongst the first in risk-adjustment, and certainly the first in European risk-adjustment.

We find that the homoscedastic smear, proposed by Duan (1983) produces significantly biased predictions. The heteroscedastic smear, proposed by Manning (1998) is an improvement, but it does not perform as well as the GLM with a logarithmic link function and a gamma variance function. The GLM offers a very flexible specification for the second part of the two-part model, and we recommend it for future applications of the two-part model.

The results of the finite mixture modelling were not so promising. We found that although the two-, three- and four-component models were superior to the one-component model in terms of goodness of fit, they each produced spurious results for some of their constituent classes. In addition, underlying causes for the generation of the mixtures could not be identified. Therefore, we found that finite mixture models were not suitable for our application of risk-adjustment in the Irish setting. Given its success in other recent health utilisation studies, such as Deb and Burgess (2002), future research could examine mixtures of alternative distributions or attempt to identify underlying causes for the results generated.

Our results highlight the importance of chronic illness, age and disability in determining prescribing expenditure. In addition, access costs reduce prescribing expenditure. Supply side effects are also noteworthy, especially the age of the GP and their prescribing style. These results are of interest, not just for risk-adjustment in the IDTS, but for risk-adjustment for GP capitation payments, which currently rely only on age, gender and access to services. In addition, the factors affecting GP

prescribing expenditure are likely to be material for other areas of the health service, such as acute care and long-term care.

We found that access adversely affected utilisation, as expected. This area merits further attention, however. It would be interesting to know what areas of the country have the highest proportions of people living far away from their GP, which might inform discussions on the appropriate geographic distribution of GPs. In addition, the role of the GP in utilisation is worthy of further research. We found that older GPs had lower levels of prescribing expenditure. Further investigation of the causes of this result would be of interest.

We found that the number of persons on the medical card in the household was negatively associated with utilisation. We also found that it was negatively correlated with chronic illness and age suggesting that it was detecting otherwise unobserved good health. We speculated that it was detecting the protective effect of living with others rather than alone and suggest that the variable be reconstructed in future work.

We found that coming from a household where the head was married was positively associated with utilisation. In addition, we found that households with a married head were negatively associated with income. This suggests that the marital status variable in our utilisation function is detecting poverty. This is worthy of further examination.

We applied a battery of tests of predictive performance and distributive implications. Of these, the examination of prediction variance, the Mincer-Zarnowitz test, the examination of the income-related distribution of need predictions and the examination of prediction error for vulnerable groups are novel in European risk-adjustment, as far as we are aware.

These tests favoured one-part over two-part models, by some distance in many cases. The Chronic Illness model and the Supply model were virtually indistinguishable, and were the best performers on most criteria. Where they differed, the Supply model was favoured. We recommend it for budget setting in the IDTS. It predicts

almost three times more variation than the NARA. It also reflects variation in actual outcomes to a greater extent, so it can better predict extremities of the distribution. It is more pro-poor and its absolute prediction error was 45% of that of the NARA for vulnerable groups, and 20% of the NARA on average. If applied it would lead to a shift of less than 5% of their NARA allocation for the middle 50% of GPs, while the middle 90% of GPs would either endure a 22% cut in their NARA allocation or enjoy a 9% increase.

An analysis of risk exposure quantified the effects of increased panel size and increased truncation on random variation in prediction error. These results may be of use to the Department of Health in negotiation with GPs on optimal risk management strategies. In the absence of information on risk preferences, we are not in a position in this study to comment on the optimal risk management strategy.

While this is predominantly a quantitative study, we examined the potential effects of the study, through two case studies. We found that the risk-adjustment exercise produced sensible results for these two GPs, which serves to validate the exercise. Furthermore, the two case studies highlighted the lack of information in general practice on the health care needs of their patients. The dissemination of chronic illness estimates based on the chronic disease score methodology would certainly offer GPs an enhancement on their current level of information and may stimulate further examination of morbidity in general practice.

The application of the Supply model would have implications for data collection on the GMS scheme. Only chronic disease scores and core variables – age, gender, distance and marital status– are collected nationally. Disability, lone parenthood and urban/rural status are available for most but not all health boards. While we have identified considerable benefits from generating these data, we are unaware of the cost of data collection, but suggest that the development of a national dataset containing the variables collected in this study should be given due consideration.

Chronic disease scores depend on the GP writing a series of particular prescriptions. Since GPs may get increases in their indicative prescribing budgets based in part on these prescriptions, a perverse incentive may arise, whereby a GP could 'generate' a

chronically ill person by writing them a series of particular prescriptions. Nevertheless, chronic disease scores and other utilisation based measures of health status are commonly used in budget setting mechanisms internationally. For instance, DRGs are used for hospital budget setting in Ireland. These too are subject to the phenomenon known as DRG-creep, where people are reclassified from one DRG to another in order to maximise the expenditure associated with them. Therefore, patterns in chronic disease scores need to be monitored in order to minimise or eliminate any perverse incentives that may arise. It is not difficult to produce ways of minimising the effects of such incentives. First, any such behaviour is clearly unethical and any GP found to be doing so should face appropriate sanctions. Second, Lamers and van Vliet (2001) indicated someone as having a chronic illness only if they have more than 181 DDDs of medicine for that chronic illness in the study year. We chose not to apply this criterion owing to concerns in some parts of the DDD field in the prescribing database. These concerns are minor and easily rectified, so this approach could be adopted. It would make gaming the formula much more difficult. Finally, a mechanism could be established whereby any person that is indicated as having, say, diabetes according to the chronic disease scores could be added to a diabetes register, once the presence of diabetes has been validated by the GP. The GP would then be expected to comply with guidelines on diabetes care and failure to do so could be linked to the GPs prescribing budget. The perverse incentive could thereby be eliminated.

Alternatively, while epidemiologically-derived estimates of health status are difficult and expensive to collect for the entire GMS population, consideration should be given to extending the chronic illness indicator used in the South Eastern Health Board nationally, either as a substitute or a complement to chronic disease scores in future formulae.

The dataset we used in this study is as rich as almost any used in risk-adjustment. It is still, however, a cross-sectional analysis. A panel dataset would allow researchers to examine the effect of health dynamics on health care need which may enrich our understanding of health care need and consequently of appropriate budget setting.

This chapter has made a number of recommendations for future research, not all of which may be possible to implement. Consequently, we attempt to prioritise these recommendations on the basis of the expected costs and benefits of each one. We place recommendations into four groups. First, there are recommendations that are possible to undertake with minor enhancements to the current dataset or methods employed and have clear expected benefits. Second, there are a number of other recommendations that are easily implemented with existing datasets and using existing methods and have clear expected benefits. Third, there are recommendations that are easily implemented but whose benefits are either more modest or more uncertain or are high cost with high expected benefits. Fourth, there is one recommendation that is expensive to implement and whose expected benefit is relatively low owing to the uncertainty in outcome.

There are three recommendations in the first group. First, perhaps the simplest recommendation to implement is to recast the numcard variable as an indicator of living alone, as we suggest above that it is detecting the protective effect of living with others rather than as a measure of large families which is in turn a measure of poverty. Second, with the collection of data on GP consultations, it would be possible to use this dataset to examine the determinants of GP consultations. This could feed directly into the GP capitation payments scheme and is of clear relevance to Irish health care policy. Third, there has been a growth of interest in the use of finite mixture models in the study of health care utilisation. In this study, we used mixtures of normal distributions. We suggested that the examination of mixtures of other distributions could prove profitable. Although analytically challenging, no new data needs to be collected and a new perspective on the results presented above may emerge.

There are two recommendations in the second group. First, further examination of indicators of prescribing style based on McGavock (1988) are easily constructed from existing datasets and could assist in our understanding of systematic differences between physicians, which have been hitherto difficult to model. Second, a comparison of the indicators of poverty generated by this study with those generated by previous research, including an examination of the role of marital status and the apparent changing status of the lone parent is of interest at least to Irish poverty

analysts and could be undertaken from existing databases and using existing techniques.

There are four recommendations in the third group. First, this dataset could also be used to further examine the role of GP age and access to services on utilisation. These recommendations are cheap to implement and are likely to be of some interest in Ireland and perhaps internationally. Second, the use of chronic disease scores as an alternative to epidemiologically-derived estimates of disease prevalence would be of interest to Irish researchers and policy-makers, since there is a chronic lack of data on community morbidity in Ireland, and may also be of interest to the international research community if they prove to be a valuable alternative. Third, quantile regression suggested that the chronic illness indicators, age and disability exhibited considerable heterogeneity and that a reconstruction of these variables may be warranted. This is simple to undertake and the results could be of considerable interest to the international research community, although the outcome of such an exercise is highly uncertain. Fourth, there is one high cost and high expected benefit recommendation. The construction of a longitudinal panel of GMS recipients would be expensive but would allow the examination of health care dynamics, with considerable benefits.

Finally, the exploration of vertical equity in the GMS is likely to be expensive with low expected benefits owing to the considerable uncertainty in outcome. Accounting for vertical equity while measuring horizontal equity poses significant conceptual difficulties which may not be possible to resolve satisfactorily. Indeed international comparisons of equity in health care such as van Doorslaer et al. (2001) generally assume that on average the health care system in each country is vertically equitable thereby assuming the problem away.

In conclusion, this study has comprehensively reviewed approaches to risk-adjustment. It has brought together a wealth of data on GMS recipients and calculated the relationship between prescribing expenditure and demographic, socio-economic, health-related, access-related and supply-side factors. It has subjected a number of competing prescribing models to a battery of tests of predictive performance and discussed the implications of such models for the management and

delivery of GP services. It has made a number of recommendations for future research. As such it should be useful for the future development of general practice in Ireland.

APPENDIX 3.1

QUALITY OF ISSUE DATE VARIABLE

This appendix considers two issues: the validity of the issue date variable in the Medical Card Register and the effect of missing issue dates on sample selection.

Inclusion criteria depend crucially on the validity of the issue date. There are two points for consideration. First, if the issue date is valid, any individual should not be reported as having any expenditure in the months prior to the issue date, as they should not be in the scheme. We took a sample of 6,000 medical cards that were issued between 16th March 2001 and 24th April 2001. We expected that these patients had no prescribing expenditure in the year prior to March. We detected 34 patients who had recorded prescribing expenditure prior to March or 0.6% of the total sample. Therefore the issue date appears to be a valid variable.

Second, issue dates refer to cards, not individuals. While we can determine the point at which the card was started, not everyone in the household may have joined the GMS scheme on the date of issue of the card, the most obvious examples being members of the household who are born after the issue date, although other household members can change in eligibility as well. While we have good data on the start date of the cardholders eligibility for GMS, we cannot be as sure about other members of the household. Data are available on individual-level eligibility, but the field in the dataset is not well coded. Nevertheless we use the information on those individual-level eligibility coded to check for differences between the issue date on the card and the issue date for individuals. A comparison of issue dates for cards and issue dates for individuals for 977 people was undertaken, as outlined in Table A3.1. The issue dates for a few cards was set at 1 January 1901, while the issue dates of individuals on these cards was recent, meaning that some differences were very large. As such, the data are skewed and medians and inter-quartile ranges are reported rather than means and standard deviations.

Table A3.1

Difference between Start Date of Medical Card and Start Date of Individuals on the Medical Card (days) by Position on Card[1]

Position	Median	25th percentile	75th percentile	N
A	13	3	21	595
B	13	4	23	158
C	15	3	25	95
D	13	3	25	61
E	14.5	0	24.5	40
Other	25	7	34,042[2]	29
Total				977

NOTE 1: Position A is the cardholder. Position B is usually the cardholders spouse. Other positions usually refer to dependants.
NOTE 2: Some cards have a default start date of 1 January 1901.

A small difference between start date on the card and start date for individuals exists in many cases, which is of no interest for our purposes of identifying people with partial-coverage for the cost year. Median difference was 13 days for cardholders and between 13 and 25 for non-cardholders. Interquartile ranges cross in all cases, so there does not appear to be any difference in the difference between start date of the card and start date of the individual, by position on the card. As such we decide to apply the card start date to all individuals on the card, not just the cardholder. The only exception we make is for those who are born during the study period, who are assigned their birthday as their start date.

As reported in the main text, the average expenditure of those individuals who did not have their issue date coded was IR£96, while it was IR£208 for those who had issue date recorded (for the period September 1999 to August 2000). Expenditures by age group are reported in Table A3.2 for the South Eastern Health Board.

Table A3.2

Mean Expenditures depending on Coding of Issue Date

Age Group	Mean expenditure with Issue date missing (IR£)	Mean expenditure with Issue date recorded (IR£)
0-4	30	36
5-11	19	30
12-15	21	31
16-24	48	65
25-34	68	130
35-44	101	151
45-54	145	217
55-64	207	307
65-69	252	350
70-74	297	369
75-79	318	385
80-85	294	396
85+	268	344
Mean	*96*	*208*

The expenditure for those with the issue date missing is always lower than those who have the issue date recorded for every age group.

We also tested for sample selection bias if those with missing issue dates are excluded. We regressed whether the issue date was reported or not on all covariates, calculated the inverse Mills' ratio and included this in a regression of expenditure on all covariates, excluding those whose issue date was missing. The inverse Mills' ratio was significant suggesting that those whose issue date was missing are not drawn randomly from the population. This suggests, *prima facie*, that excluding those whose issue date is not recorded would be both inefficient and biased. However this test for sample selection bias is predicated on the assumption that those with missing issue dates have a full year's expenditure data. However, it is reasonable to assume that those with missing issue dates do not have a full year's expenditure data, as described in the main text. If we accept that they do not have a full year's expenditure data, then we cannot test for sample selection bias and the effect of excluding those with missing issue dates becomes unclear.

In conclusion, the issue date appears to be valid, while the effect of excluding those observations that have are missing the issue date cannot be tested using standard sample selection criteria, as we suspect that they received their cards recently and therefore do not qualify for a full year's expenditures.

APPENDIX 3.2
DRUGS INCLUDED IN PRESCRIBING STYLE INDICATORS AND CHRONIC ILLNESS INDICATORS

Table A3.3 describes the drugs included in each of three indicators of prescribing style, based on McGavock (1988). Meanwhile, Table A3.4 outlines the link between the original chronic disease scores and the indicators used in our study. There were 28 chronic conditions identified in Lamers (1999a). A number of drugs indicates as identifiers of a chronic condition are categorised as 'often presumptive' or 'symptomatic' in McGavock's classification. We wish only to indicate drugs as identifiers of a chronic condition if they are 'specific', that is, if they follow from an accurate diagnosis. Those drugs that are excluded on this basis are marked in italics in Table A3.4. In addition, in order to keep our model as parsimonious as possible, a number of similar chronic diseases are grouped together, and drugs rarely prescribed in General Practice are excluded entirely. Any chronic illness that had a prevalence rate of less than 1% was assigned to the latter category.

Table A3.3

Teleological Classification of Main Therapeutic Areas reflecting Perceived Use in General Practitioner

Drug Type	Therapeutic Area	ATC Code
Specific	Antiepileptics	N03
	Anti-Parkinson drugs	N04
	Antidepressants	N06A
	Heart preparations	C01, C07, C08, C09, C10
	Diuretics	C03
	Antihypertensives	C02
	Asthma preparations	R03
	Hypoglycaemic agents	A10
	Corticosteriods	H02
	Thyroid/Anti-thyroid agents	H03
	Anticoagulants	B01
	Vaginal preparations	G01
	Eye preparations	S01
	Dressings	V20
	Combinations for eradication of Helicobacter pylori	A02BD
	Antipsoriatics	D05
	Anti-acne preparations	D10
	Antimycobacterials	J04
	Antimigraine preparations	N02C
Symptomatic	Hypnotics	N05C
	Analgesics minor	N02 excl. N02C
	Antacids	A02 excl. A02B, A03, A04
	Laxatives	A06
	Expectorant/cough suppressants	R05
	Anti-inflammatory (rheumatism)	M01, M02, M03
	Vitamins	A11
	Antihistamines	R06
Often Presumptive	Sedatives/tranquillisers	N05B
	Proton pump inhibitors, H_2 receptor antagonists etc.	A02B excl. A02BD
	Vasodilators/vasoconstrictors	C04
	Penicillins and other anti-microbials	J01, J02
	Topical skin preparations	D01, D02, D06, D07

Source: McGavock (1988: 192) (modified)

Table A3.4

ATC-codes for chronic conditions

Chronic Condition	Chronic Disease Score category	ATC-code	Description of ATC-code
Cardio-vascular disease	Coronary and peripheral vascular disease	B01A, *C04AD03*	Antithrombotic agents, *pentoxifylline*
	Hypertension	C02	Antihypertensives: Antiadrenergic agents, centrally acting, ganglion-blocking , peripherally acting, other antihypertensives
		C03A, C03EA01	Low-ceiling diuretics, thiaziden
		C07	Beta blocking agents
		C08	Calcium channel blockers
		C09A , C09B	Angiotensin-converting enzyme (ACE) inhibitors
	Hyperlipidemia	C10A	Cholesterol and triglyceride reducers
	Cardiac disease / ASCVD / CHF	C01	Cardiac therapy: Cardiac stimulants and glycosides Antirrhythmics, class I and III Vasodilators used in cardiac diseases
		C03C, C03EB01	High-ceiling diuretics
Epilepsy	Epilepsy	N03A (excluding N03AE01)	Antiepileptics
Rheumatologic conditions, pain and inflammation	Rheumatologic conditions	H02	Corticosteroids for systematic use
		M01CB, M01CC01, P01BA02, L01BA01, A07EC01	*Gold preparations, Penicillamine,* Hydroxychloroquine, Methotrexate, Sulfasalazine
	Pain and inflammation	*M01A*	*Non-steroids antiinflammatory and antirheumatic products*
	Pain	*N02A*	*Opioids*
Diabetes	Diabetes	A10A	Insulins
		A10B	Oral blood glucose lowering drugs
Glaucoma	Glaucoma	S01E	Antiglaucoma preparations
Acid peptic disease	Acid peptic disease	*A02A, A02B* (A02BD included only)	*Antacids. drugs for treatment of peptic ulcer* (Combinations for eradication of helicobacter pylori included only)
Respiratory Illness, asthma	Respiratory illness, asthma	R03	Anti-asthmatics

Thyroid	Thyroid disorders	H03A, H03B	Thyroid preparations, Antithyroid preparations
Psychiatric Illness	Depression	N06AA	Tricyclic antidepressants
		N06AB, N06AE	Selective serotonin reuptake inhibitors
		N06AF, N06AG	Monoamine oxidase inhibitors
		N06AX	Other antidepressants
	Psychotic illness (including bipolar disorders)	N05A	Antipsychotics
	Anxiety and tension	*N05B*	*Anxiolytics*
Rarely Prescribed in General Practice	Tuberculosis	J04A	Drugs for treatment of tuberculosis
	Malignancies	L01 (excluding L01BA01) , L03AA02 / 03 / 10, A04AA	Antineoplastic agents, Filgrastim, molgramostim, lenograstim, Serotonin (5HT3) antagonists
	Parkinson's disease	N04B	Dopaminergic agents
	Renal disease (including ESRD)	B03XA01, V03AE01	Erythropoietin, polystyrene sulphonate
	Cystic fibrosis	A09AA02	Multienzymes
	Transplantations	L04AA01 / 05 / 06, L04AX01	Ciclosporin, Tacrolimus, mycophenolic acid, Azathioprine
	Gout	M04A	Antigout preparations
	Crohn's and ulcerative colitis	A07EC (excluding A07EC01)	Mesalazine, olsalazine

NOTES: ASCVD: arteriosclcrotic cardiovascular disease, CHF: congestive heart failure, ESRD: end stage renal disease, HIV: human immunodeficiency virus

Source: Lamers, 2000 (personal communication).

APPENDIX 3.3

INSTRUMENTAL VARIABLES FOR CHRONIC ILLNESS INDICATORS

Table A3.5 to A3.12 describe the first stage of a two stage least squares approach to overcoming measurement error in the chronic illness variables.

Table A3.5

Predictors of Cardiovascular Disease

Variable	Prob.*	P>|z|
Age	0.008	0.000
Agesq	0.000	0.000
Gender	0.005	0.000
Marital	0.011	0.000
Disabil	0.016	0.000
Lonepare	-0.005	0.045
Rural	0.000	0.761
Numcard	-0.007	0.000
3-5 miles	-0.001	0.309
5-7 miles	0.001	0.337
7-10 miles	-0.006	0.001
10+ miles	0.001	0.650
NEHB	-0.007	0.000
SEHB	-0.003	0.003
GPage	-0.001	0.173
GPagesq	0.000	0.385
Nurse	0.003	0.000
Sec	0.000	0.748
RPA	-0.014	0.000
Decpanel	0.000	0.724
Specific	0.030	0.000
Symptomatic	0.027	0.000
Presum	0.022	0.000
Pseudo-R^2		0.2949
P>χ^2		0.00
N		152,966

*Probabilities of a positive outcome at the mean for that variable are reported rather than coefficients, to aid interpretation

The R^2 of almost 30% is reasonably high. Cardiovascular disease is best predicted by age and the percentage of the GP's prescribing that is specific, symptomatic and often presumptive. Other significant variables affect probability by less than 2% at the mean. As a group of variables, distance to the GP is unimportant, highlighting the contention in the main text that access to GP services is unlikely to generate measurement error in the chronic illness variables.

Table A3.6

Predictors of Epilepsy

Variable	Prob.*	P>|z|
Age	0.001	0.000
Agesq	0.000	0.000
Gender	-0.002	0.000
Marital	-0.004	0.000
Disabil	0.014	0.000
Lonepare	-0.007	0.000
Rural	0.001	0.051
Numcard	-0.003	0.000
3-5 miles	0.001	0.259
5-7 miles	0.000	0.557
7-10 miles	0.000	0.808
10+ miles	0.002	0.326
NEHB	0.001	0.233
SEHB	0.002	0.006
GPage	0.000	0.407
GPagesq	0.000	0.633
Nurse	0.001	0.071
Sec	-0.001	0.265
RPA	-0.002	0.001
Decpanel	0.000	0.385
Specific	0.000	0.000
Symptomatic	0.000	0.047
Presum	0.000	0.235
Pseudo-R^2		0.0831
P>χ^2		0.00
N		152966

*Probabilities of a positive outcome at the mean for that variable are reported rather than coefficients, to aid interpretation

The epilepsy model has a poor level of explained variance at only 8%, illustrating the problem of weak instruments. Meanwhile only disability affects the probability of being coded as having epilepsy by more than 1%.

Table A3.7

Predictors of Rheumatology

Variable	Prob.*	P>|z|
Age	0.001	0.000
Agesq	0.000	0.000
Gender	-0.004	0.000
Marital	0.004	0.000
Disabil	0.007	0.000
Lonepare	-0.002	0.286
Rural	0.001	0.170
Numcard	-0.003	0.000
3-5 miles	-0.001	0.343
5-7 miles	0.000	0.994
7-10 miles	-0.002	0.213
10+ miles	-0.001	0.504
NEHB	-0.006	0.000
SEHB	-0.002	0.000
GPage	0.000	0.429
GPagesq	0.000	0.378
Nurse	0.002	0.001
Sec	0.000	0.917
RPA	-0.003	0.000
decpanel	0.000	0.944
Specific	0.001	0.000
Symptomatic	0.001	0.000
Presum	0.000	0.021
Pseudo-R^2		0.096
$P>\chi^2$		0.00
N		152966

*Probabilities of a positive outcome at the mean for that variable are reported rather than coefficients, to aid interpretation

The R^2 of almost 10% is again quite low, while no variable affects the probability of success by as much as 1%.

Table A3.8

Predictors of Diabetes

Variable	Prob.*	P>|z|
Age	0.001	0.000
Agesq	0.000	0.000
Gender	-0.003	0.000
Marital	0.003	0.000
Disabil	0.004	0.000
Lonepare	-0.002	0.190
Rural	-0.001	0.041
Numcard	-0.001	0.000
3-5 miles	0.000	0.584
5-7 miles	0.000	0.420
7-10 miles	0.000	0.848
10+ miles	0.000	0.901
NEHB	-0.001	0.234
SEHB	0.002	0.000
GPage	0.000	0.173
GPagesq	0.000	0.203
Nurse	0.001	0.219
Sec	0.001	0.228
RPA	-0.002	0.005
Decpanel	0.000	0.126
Specific	0.001	0.000
Symptomatic	0.000	0.030
Presum	0.001	0.001
Pseudo-R^2		0.1054
$P>\chi^2$		0.00
N		152966

*Probabilities of a positive outcome at the mean for that variable are reported rather than coefficients, to aid interpretation

The diabetes model has an R^2 of almost 11%. Although there are a number of variables that are statistically significant, none have a large effect on the probability of being coded as suffering from diabetes.

Table A3.9

Predictors of Respiratory Illness

Variable	Prob.*	P>|z|
Age	0.000	0.004
Agesq	0.000	0.000
Gender	-0.013	0.000
Marital	0.010	0.000
Disabil	0.011	0.000
Lonepare	0.003	0.250
Rural	-0.005	0.001
Numcard	-0.006	0.000
3-5 miles	-0.002	0.214
5-7 miles	-0.006	0.000
7-10 miles	-0.010	0.001
10+ miles	0.002	0.595
NEHB	-0.010	0.000
SEHB	-0.002	0.219
GPage	0.000	0.604
GPagesq	0.000	0.530
Nurse	0.004	0.001
Sec	-0.001	0.652
RPA	-0.014	0.000
Decpanel	0.000	0.068
Specific	0.002	0.000
Symptomatic	0.002	0.000
Presum	0.001	0.000
Pseudo-R^2		0.0328
$P>\chi^2$		0.00
N		152966

*Probabilities of a positive outcome at the mean for that variable are reported rather than coefficients, to aid interpretation

With an R^2 of only 3% the respiratory model has the lowest explanatory power, indicating its weakness as an instrumental variable. A number of variables affect probability of being coded as a sufferer of respiratory illness by more than 1%, including gender, marital status, disability, living 7-10 miles from the GP, being a resident in the North East Health Board and attending a GP in receipt of rural practice allowance.

Table A3.10

Predictors of Glaucoma

Variable	Prob.*	P>|z|
Age	0.000	0.000
Agesq	0.000	0.040
Gender	-0.001	0.000
Marital	0.001	0.000
Disabil	0.000	0.277
Lonepare	0.000	0.334
Rural	0.000	0.008
Numcard	0.000	0.000
3-5 miles	0.000	0.638
5-7 miles	0.000	0.586
7-10 miles	0.000	0.452
10+ miles	0.000	0.928
NEHB	-0.001	0.000
SEHB	0.000	0.635
GPage	0.000	0.489
GPagesq	0.000	0.450
Nurse	0.000	0.140
Sec	0.000	0.692
RPA	0.000	0.006
Decpanel	0.000	0.250
Specific	0.000	0.000
Symptomatic	0.000	0.006
Presum	0.000	0.051
Pseudo-R^2		0.1692
$P>\chi^2$		0.00
N		152966

*Probabilities of a positive outcome at the mean for that variable are reported rather than coefficients, to aid interpretation

Although the glaucoma model has an R^2 of almost 17%, no variable stands out as a predictor of glaucoma.

Table A3.11

Predictors of Thyroid Illness

Variable	Prob.*	P>|z|
Age	0.001	0.000
Agesq	0.000	0.000
Gender	0.012	0.000
Marital	0.000	0.355
Disabil	0.003	0.000
Lonepare	-0.001	0.244
Rural	0.000	0.831
Numcard	-0.001	0.000
3-5 miles	-0.001	0.252
5-7 miles	0.000	0.609
7-10 miles	0.001	0.146
10+ miles	0.000	0.713
NEHB	0.000	0.933
SEHB	0.002	0.000
GPage	0.000	0.852
GPagesq	0.000	0.893
Nurse	0.001	0.017
Sec	0.000	0.992
RPA	-0.002	0.001
Decpanel	0.000	0.259
Specific	0.000	0.000
Symptomatic	0.000	0.008
Presum	0.000	0.087
Pseudo-R^2		0.1347
$P>\chi^2$		0.00
N		152966

*Probabilities of a positive outcome at the mean for that variable are reported rather than coefficients, to aid interpretation

Similar to the glaucoma model, the thyroid model has a higher level of explained variance that many of the other models presented above, but the effect of each statistically significant variable is very slight.

Table A3.12

Predictors of Psychiatric Illness

Variable	Prob.*	P>|z|
Age	0.005	0.000
Agesq	0.000	0.000
Gender	0.014	0.000
Marital	-0.016	0.000
Disabil	0.038	0.000
Lonepare	-0.010	0.000
Rural	0.000	0.869
Numcard	-0.008	0.000
3-5 miles	-0.003	0.048
5-7 miles	-0.001	0.224
7-10 miles	-0.007	0.002
10+ miles	0.003	0.437
NEHB	-0.005	0.000
SEHB	0.002	0.101
GPage	0.001	0.083
GPagesq	0.000	0.036
Nurse	0.002	0.026
Sec	-0.003	0.067
RPA	-0.011	0.000
decpanel	0.000	0.564
specific	0.000	0.173
symptomatic	-0.000	0.714
presum	-0.000	0.153
Pseudo-R^2		0.1209
$P>\chi^2$		0.00
N		152966

*Probabilities of a positive outcome at the mean for that variable are reported rather than coefficients, to aid interpretation

The R^2 for the psychiatric model is just over 12%, while those predictors that affect the probability of being coded as a sufferer of psychiatric illness by more than 1% are gender, marital status, lone parenthood, disability and attending a GP in receipt of rural practice allowance. In addition age is an important predictor.

As described in the main text, the explanatory power of the first stage of two stage least squares is low, so alternative ways of testing for measurement error must be sought.

APPENDIX 3.4

PREDICTING LOGARITHM OF INCOME

The main text uses a linear model of predicted income to calculate the current income-related distribution of prescribing expenditure and the distributional effects of competing utilisation models. This appendix outlines the results of a model of income that uses the natural logarithmic transformation of the response variable. As outlined in the main text, a number of people had non-positive income in the Household Budget Survey. We assume that their income was IR£0.01 for the purposes of logarithmic transformation.

Table A3.13

Factors affecting log-transformed household income for those households where all members are GMS recipients

Variable	Coef.	P|t|>0
M06_13	0.082	0.00
M14_20	0.210	0.00
M21_44	0.481	0.00
M45_64	0.348	0.00
M65etc	0.429	0.00
F00_05	0.110	0.00
F06_13	0.086	0.00
F14_20	0.274	0.00
F21_44	0.525	0.00
F45_64	0.395	0.00
F65etc	0.383	0.00
Marital	0.064	0.04
Incaporill	-0.101	0.01
Lonepare	0.248	0.00
_cons	4.489	0.00
R^2	0.4487	
Prob >F	0.0000	
N	2,247	

*Heteroscedasticity robust standard errors

Source: Household Budget Survey, 1999/2000 (2002)

The R^2 of 0.45 is slightly higher than that of 0.44 for the linear model, as reported in the main text. The reference category is again males aged 0 to 5. As with the untransformed model, the main factors affecting household income are the number of people aged 21 to 64. Unlike in the linear model, marital status is positive and significant, although its effect is the smallest of all variables included in the model.

Table A3.14 describes the factors affecting the natural logarithm of income for households where some but not all members are GMS recipients.

Table A3.14

Factors affecting log-transformed household income for households where some but not all members are GMS recipients.

Variable	Coef.	P|t|>0*
M06_13	0.048	0.08
M14_20	0.140	0.00
M21_44	0.366	0.00
M45_64	0.273	0.00
M65etc	0.122	0.01
F00_05	0.101	0.06
F06_13	0.042	0.14
F14_20	0.135	0.00
F21_44	0.295	0.00
F45_64	0.256	0.00
F65etc	0.240	0.00
Marital	0.003	0.94
Incaporill	-0.186	0.00
Lonepare	-0.280	0.16
_cons	5.446	0.00
R^2	0.3864	
Prob >F	0.0000	
N	772	

*Heteroscedasticity robust standard errors
Source: Household Budget Survey, 1999/2000 (2002)

The model has a lower level of explained variance than the linear model reported in the main text (R^2 of 38.64% v. 40.87%). The reference case is again males 0 to 5. As with the linear model, the number of males aged 21 to 64 and the number of females aged over 20 have large effects on household income. As with the linear model, lone parenthood and marital status are insignificantly related to household income and number of household members with an incapacity or illness is negatively related to income.

Given that the explanatory power of the log-transformed model is lower than the linear model for the second equation and that it is only marginally higher for the first equation, and given that the linear model is easier to interpret, the linear model was preferred.

APPENDIX 3.5

BUDGET NEUTRAL DRUGS

Table A3.15

Description of Budget Neutral Drugs for 2000

ATC code	Brand Name	Generic Name	Description
A04AA01	Zofran	Ondansetron	Serotonin antagonists used to prevent nausea after chemotherapy
A04AA02	Kytril	Granisetron	Prevent nausea after chemotherapy
A09AA02	Creon, nutrizym, pancrease	Multi-enzymes	Lipase, protease, etc. Used to help digestion when pancreas not working properly
J01DA11	Fortum	Ceftazidime	Cephlasporins and related substances (a beta-lactam antibacterial)
J01XA02	Tarcocid	Teicoplanin	Glycopeptide antibacterials
L02AE01	Supercur	Buserelin	Gonadotropin releasing hormone analogues. Hormone agent used to treat prostrate cancer
L02BG03	Arimidex	Anastrozole	Enyzme inhibitors. Used to treat advanced breast cancer in post menopausal women.
L04AX01	Imuran	Azathioprine	Immunosuppressive agent, used to prevent rejection of kidney transplants. Also used for severe rheumatoid arthritis.
V06CA81	Aminogran, Lofenalac, Maxamum, Maxamaid	Not named	Nutrients without phenylalanine (for infants)

Source: Southern Health Board

In addition to this list, a provision is made for the prescription of sildenafil (brand name Viagra) which essentially makes it budget neutral. Sildenafil is used to treat erectly dysfunction in males. Also lipid-lowering agents (ATC code C10AA) were made budget neutral during 2000.

APPENDIX 5.1

ESTIMATION OF MODELS WITH DIFFERENT TREATMENT OF OUTLIERS

Table A5.1

NARA with different treatment of outliers

Variables	All		IDTS		This Study	
	Coef.	P>|t|	Coef.	P>|t|	Coef.	P>|t|
constant	295.43	0.00	254.63	0.00	249.16	0.00
age04	-261.00	0.00	-224.03	0.00	-219.53	0.00
age515	-265.78	0.00	-228.77	0.00	-223.87	0.00
age1644	-176.38	0.00	-147.16	0.00	-144.40	0.00
age6569	91.41	0.00	78.11	0.00	76.92	0.00
age70	88.43	0.00	96.86	0.00	95.84	0.00
N	400,751		400,751		400,751	
RMSE	428.25		334.87		315.3	
$\overline{R}^2$ (%)		9.66		12.29		13.23

Table A5.2

Demographic model with different treatment of outliers

Variables	All		IDTS		This Study	
	Coef.	P>|t|	Coef.	P>|t|	Coef.	P>|t|
constant	42.35	0.00	37.77	0.00	35.40	0.00
Age	4.55	0.00	3.64	0.00	3.56	0.00
Agesq	0.00	0.61	0.01	0.00	0.01	0.00
Gender	-5.94	0.00	-2.37	0.05	-1.38	0.17
Marital	25.55	0.00	16.69	0.00	16.67	0.00
Disabil	127.92	0.00	105.45	0.00	100.22	0.00
Lonepare	-2.50	0.46	-1.33	0.58	-0.45	0.87
Numcard	-12.79	0.00	-11.25	0.00	-14.81	0.00
Rural	-19.77	0.00	-15.54	0.00	-11.15	0.00
3-5 miles	-9.53	0.00	-8.06	0.00	-7.87	0.00
5-7 miles	-10.98	0.00	-9.45	0.00	-9.18	0.00
7-10 miles	-32.31	0.00	-26.59	0.00	-25.35	0.00
10+ miles	-25.01	0.01	-23.23	0.00	-22.88	0.00
NEHB	10.83	0.06	7.54	0.13	6.29	0.00
SEHB	11.53	0.05	7.50	0.13	6.64	0.00
N	400,751		400,751		400,751	
RMSE	425.52		332.30		312.82	
$\overline{R}^2$ (%)		10.81		13.96		14.59

Table A5.3

Chronic Illness model with different treatment of outliers

Variables	All		IDTS		This Study	
	Coef.	P>|t|	Coef.	P>|t|	Coef.	P>|t|
constant	13.38	0.00	11.48	0.00	10.11	0.00
Age	3.76	0.00	2.86	0.00	2.81	0.00
Agesq	-0.03	0.00	-0.02	0.00	-0.01	0.00
Gender	-5.41	0.00	-1.91	0.07	-1.12	0.24
Marital	23.65	0.00	16.24	0.00	16.03	0.00
Disabil	68.46	0.00	52.79	0.00	49.59	0.00
Lonepare	1.47	0.57	2.83	0.14	3.59	0.10
Numcard	-7.97	0.01	-7.10	0.00	-8.90	0.00
Rural	-12.75	0.00	-9.44	0.00	-7.11	0.00
3-5 miles	-4.96	0.04	-4.15	0.03	-4.17	0.00
5-7 miles	-9.32	0.00	-7.85	0.00	-7.65	0.00
7-10 miles	-21.31	0.00	-16.95	0.00	-15.99	0.00
10+ miles	-20.68	0.00	-16.58	0.01	-16.20	0.00
NEHB	28.59	0.00	23.35	0.00	21.55	0.00
SEHB	16.43	0.00	11.72	0.00	10.78	0.00
CVD	322.81	0.00	262.63	0.00	261.45	0.00
Epi	333.61	0.00	287.85	0.00	274.63	0.00
Rheum	376.62	0.00	327.56	0.00	316.39	0.00
Diabetes	439.63	0.00	392.85	0.00	383.20	0.00
Glau	265.39	0.00	270.04	0.00	270.93	0.00
Respir	372.65	0.00	344.06	0.00	331.69	0.00
Thyroid	103.92	0.00	101.43	0.00	107.09	0.00
Psych	395.87	0.00	367.33	0.00	354.93	0.00
comor	-42.91	0.00	-54.42	0.00	-64.06	0.00
N	362,072		362,072		362,072	
RMSE	380.24		288.60		269.70	
$\overline{R}^2$ (%)		31.02		37.17		38.75

Table A5.4

Supply model with different treatment of outliers

Variables	All		IDTS		This Study	
	Coef.	P>|t|	Coef.	P>|t|	Coef.	P>|t|
constant	33.52	0.73	-26.27	0.75	-42.10	0.13
Age	3.78	0.00	2.86	0.00	2.81	0.00
Agesq	-0.03	0.00	-0.01		-0.01	0.00
Gender	-5.47	0.00	-2.04	0.08	-1.22	0.24
Marital	23.72	0.00	16.30	0.00	16.05	0.00
Disabil	66.37	0.00	50.21	0.00	47.21	0.00
Lonepare	1.00	0.72	2.77	0.17	3.48	0.13
Numcard	-1.52	0.61	0.25	0.92	-8.77	0.00
Rural	-12.64	0.00	-9.31	0.00	0.09	0.94
3-5 miles	-4.09	0.13	-3.83	0.06	-3.91	0.01
5-7 miles	-8.63	0.00	-7.10	0.00	-6.85	0.00
7-10 miles	-17.86	0.00	-13.55	0.00	-12.56	0.00
10+ miles	-15.07	0.04	-10.52	0.09	-10.08	0.02
NEHB	28.69	0.00	23.92	0.00	22.11	0.00
SEHB	21.59	0.00	17.02	0.00	16.07	0.00
CVD	320.37	0.00	261.00	0.00	259.83	0.00
Epi	332.55	0.00	287.66	0.00	274.79	0.00
Rheum	372.93	0.00	322.86	0.00	312.19	0.00
Diabetes	434.69	0.00	389.11	0.00	380.44	0.00
Glau	265.53	0.00	268.90	0.00	269.69	0.00
Respir	367.90	0.00	340.90	0.00	328.84	0.00
Thyroid	102.55	0.00	101.86	0.00	107.32	0.00
Psych	394.17	0.00	365.07	0.00	352.62	0.00
comor	-40.43	0.00	-52.77	0.00	-62.50	0.00
GPage	-8.91	0.00	-7.04	0.00	-6.64	0.00
GPagesq	0.09	0.00	0.07	0.00	0.06	0.00
Nurse	9.34	0.03	9.15	0.01	8.92	0.00
Sec	-5.03	0.37	-4.26	0.40	-4.38	0.02
RPA	-1.86	0.79	-5.12	0.39	-5.22	0.00
Decpanel	0.00	0.66	0.00	0.49	0.00	0.03
Specific	140.62	0.07	142.11	0.04	149.73	0.00
symptomatic	422.27	0.00	431.11	0.00	427.74	0.00
Presum	299.54	0.01	339.10	0.00	348.51	0.00
N	305,887		305,887		305,887	
RMSE	381.50		288.55		269.71	
$\overline{R}^2$ (%)		30.96		37.31		38.91

Table A5.5

Logarithmic model of Users with different treatment of outliers

Variables	All		IDTS		This Study	
	Coef.	P>|t|	Coef.	P>|t|	Coef.	P>|t|
constant	2.173	0.000	2.066	0.000	2.064	0.00
Age	0.045	0.000	0.042	0.000	0.041	0.00
Agesq	0.000	0.000	0.000	0.000	-0.000	0.00
Gender	0.073	0.000	0.091	0.000	0.093	0.00
Marital	0.078	0.000	0.052	0.000	0.061	0.05
Disabil	0.226	0.000	0.219	0.000	0.215	0.00
Lonepare	0.033	0.004	0.032	0.005	0.032	0.00
Numcard	0.006	0.662	0.010	0.485	0.010	0.17
Rural	-0.085	0.000	-0.078	0.000	-0.079	0.00
3-5 miles	-0.044	0.001	-0.040	0.002	-0.041	0.00
5-7 miles	-0.047	0.000	-0.043	0.000	-0.043	0.00
7-10 miles	-0.079	0.001	-0.069	0.002	-0.069	0.00
10+ miles	-0.080	0.025	-0.070	0.049	-0.070	0.00
NEHB	0.064	0.008	0.070	0.003	0.071	0.00
SEHB	0.024	0.292	0.022	0.322	0.024	0.01
CVD	1.488	0.000	1.366	0.000	1.365	0.00
Epi	1.541	0.000	1.538	0.000	1.533	0.00
Rheum	1.578	0.000	1.556	0.000	1.551	0.00
Diabetes	1.799	0.000	1.788	0.000	1.783	0.00
Glau	1.488	0.000	1.507	0.000	1.508	0.00
Respir	1.698	0.000	1.717	0.000	1.710	0.00
Thyroid	1.018	0.000	0.982	0.000	0.985	0.00
Psych	1.655	0.000	1.671	0.000	1.665	0.00
comor	-1.084	0.000	-1.059	0.000	-1.065	0.00
GPage	-0.047	0.000	-0.046	0.000	-0.045	0.00
GPagesq	0.000	0.000	0.000	0.000	0.000	0.00
Nurse	0.054	0.019	0.054	0.020	0.054	0.00
Sec	-0.041	0.188	-0.040	0.212	-0.040	0.00
RPA	-0.033	0.380	-0.042	0.252	-0.043	0.00
Decpanel	0.000	0.593	0.000	0.557	0.000	0.03
Specific	1.071	0.028	1.107	0.027	1.118	0.00
symptomatic	2.622	0.000	2.815	0.000	2.817	0.00
Presum	3.049	0.000	3.171	0.000	3.180	0.00
N	236,308		236,308		236,308	
$\overline{R}^2$ (%)	48.93		47.23		47.24	

Table A5.6

Generalised Linear Model of Users with different treatment of outliers

Variables	All		IDTS		This Study	
	Coef.	P>|t|	Coef.	P>|t|	Coef.	P>|t|
constant	4.65	0.00	4.11	0.00	3.978	0.00
Age	0.04	0.00	0.04	0.00	0.044	0.00
Agesq	0.00	0.00	0.00	0.00	-0.000	0.00
Gender	-0.04	0.00	-0.02	0.05	-0.015	0.09
Marital	0.03	0.11	0.01	0.34	0.014	0.31
Disabil	0.27	0.00	0.24	0.00	0.233	0.00
Lonepare	-0.18	0.00	-0.14	0.00	-0.126	0.00
Numcard	-0.02	0.26	-0.01	0.51	-0.009	0.52
Rural	-0.12	0.00	-0.10	0.00	-0.100	0.00
3-5 miles	0.00	0.94	0.00	0.93	-0.002	0.88
5-7 miles	-0.03	0.04	-0.02	0.06	-0.024	0.06
7-10 miles	-0.03	0.45	-0.03	0.24	-0.027	0.30
10+ miles	-0.08	0.06	-0.06	0.15	-0.055	0.15
NEHB	0.10	0.00	0.09	0.00	0.085	0.00
SEHB	0.06	0.01	0.04	0.06	0.033	0.08
CVD	0.92	0.00	0.83	0.00	0.828	0.00
Epi	1.34	0.00	1.31	0.00	1.284	0.00
Rheum	1.20	0.00	1.16	0.00	1.147	0.00
Diabetes	1.35	0.00	1.35	0.00	1.344	0.00
Glau	0.94	0.00	0.97	0.00	0.975	0.00
Respir	1.25	0.00	1.26	0.00	1.247	0.00
Thyroid	0.70	0.00	0.69	0.00	0.689	0.00
Psych	1.25	0.00	1.26	0.00	1.244	0.00
comor	-0.76	0.00	-0.75	0.00	-0.759	0.00
GPage	-0.03	0.01	-0.03	0.00	-0.032	0.00
GPagesq	0.00	0.01	0.00	0.00	0.000	0.00
Nurse	0.05	0.04	0.06	0.01	0.056	0.01
Sec	-0.09	0.01	-0.07	0.01	-0.069	0.02
RPA	0.01	0.69	-0.01	0.84	-0.007	0.83
Decpanel	0.00	0.47	0.00	0.21	0.000	0.19
Specific	-0.34	0.44	-0.02	0.96	0.057	0.89
symptomatic	0.92	0.08	1.28	0.01	1.311	0.01
Presum	-0.04	0.94	0.84	0.13	1.000	0.06
N	236,308		236,308		236,308	
Log Likelihood	-1,459,995		-1,435,739		-1,432,162	

APPENDIX 6.1

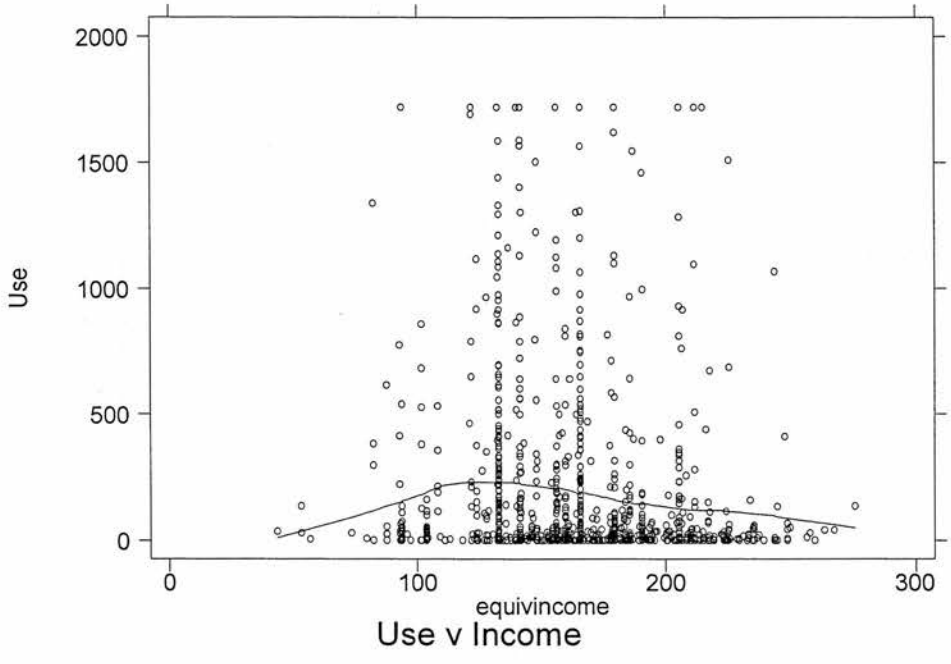
AN EXPLORATION OF UNMET NEED

A6.1 INTRODUCTION

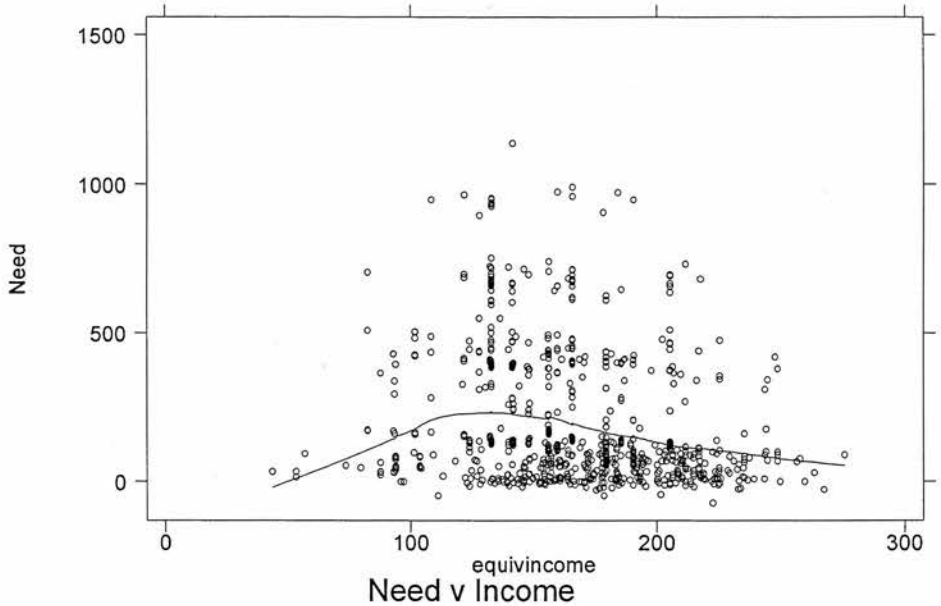
This appendix tests for the existence of income-related unmet need in the GMS. Figure A6.1 illustrates the relationship between utilisation and income in the first graph, and need and income in the second graph, using locally-weighted regression. Need is defined as the capitation predictions from the Supply model. If health status is measured accurately, such that an income variable is not detecting an element of health status, and if there is equal utilisation for equal need, then there should be no relationship between income and utilisation. To the extent that income is partially detecting health status, the relationship is expected to be negative. By extension, there should be no relationship between need and income, unless health status is measured with error, in which case the relationship is expected to be negative.

Locally-weighted regression produces a smoothed value for each observation (x_i,y_i) by running a regression of y on x using only the observation (x_i,y_i) and a number of observations near that point, with greater downweighting of observations the further away from (x_i,y_i) that they are. The number of observations that are included in each locally-weighted regression depends on the bandwidth applied. We apply a bandwidth of 0.8, meaning that the 80% of observations closest to (x_i,y_i) are included in the locally-weighted regression that produces its smoothed value (StataCorp, 2000).

Figure A6.1 Relationship between Utilisation and Income and Need and Income



Use v Income



Need v Income

Similar to Sutton and Lock (2000), we find that there is a negative relationship between use and income for the majority of people, but for the poorest group the relationship is actually positive, meaning that the very poorest have lower utilisation than the next poorest. Consequently, 'need' (or predicted utilisation controlling for

supply-side variables and access) has a similar relationship with income, where the poorest have lower 'need' than higher income groups.

A6.2 STATISTICAL ANALYSIS

A6.2.1 Spline Regression

The degree of variation in health care utilisation is very apparent from Figure A6.1, meaning that the patterns that emerge may not be statistically significant. Following Lock and Sutton (2000) we test for unmet need by examining if the relationship between income and utilisation is constant across all levels of income. Since the changes in slope in both graphs in Figure A6.1 are between the 10th and 20th decile, we apply spline regression with a knot at each decile. We also apply a spline regression with a knot at each quartile, in case the decile-level spline regression does not produce a consistent pattern. The knots can be expressed as dummy variables, K_j, where $K_j = 1$ if $income_i \geq t_j$ or 0 otherwise, $income_i$ is the income of individual i and t_j vary from the 10th decile to the 90th decile. We can express the spline specification as:

$$y_i = \beta_0 + \beta_1 income_i + \sum_{j=1}^{J-1} K_j \lambda_j v_i + \varepsilon_i$$

where y_i is prescribing expenditure of i, ε_i is a independent, identically distributed error term, β, and λ_j are coefficient vectors and $v_i = income_i - t_j$.

Therefore for those with income less than the 10th percentile (first decile), income is represented as $income_i$ only. For those with income in the second decile, income is represented by $income_i$ and their income less the income at the 10th decile. For those with income in the third decile, income is represented by $income_i$, their income less the income at the 10th decile and their income less the income at the 20th decile. The process continues for remaining incomes.

The results of the spline regressions at decile-level and quartile-level are reported in Table A6.1.

Table A6.1

Relationship between Utilisation and Income by level of Income

Variable	Coef.	P>|t|	Variable	Coef.	P>|t|
Decile 1	1.23	0.00	Quartile 1	2.75	0.00
Decile 2	9.44	0.00	Quartile 2	-5.15	0.00
Decile 3	-19.18	0.00	Quartile 3	-1.36	0.00
Decile 4	1.20	0.06	Quartile 4	3.02	0.00
Decile 5	22.98	0.00			
Decile 6	-57.62	0.00			
Decile 7	51.14	0.00			
Decile 8	-14.44	0.00			
Decile 9	6.43	0.00			
Decile 10	-3.03	0.00			
Constant	54.86	0.00	Constant	-82.51	0.00
R^2 (%)		5.43			3.91
N		400,751			400,751

Although a pattern appears to emerge from Figure A6.1, the degree of variation in health care utilisation means that this pattern is not borne out by the regression analysis. We find that the relationship between utilisation and income changes with every decile except the fourth one. These results report the marginal effect of introducing a new spline with each decile, as opposed to the absolute effect. For instance, the absolute slope of the relationship between utilisation and income is +10.67 at the second decile (1.23 + 9.44). Therefore, not only do marginal effects vary in sign as reported here, but the absolute effect changes sign by decile. The third, fourth, sixth, eight and tenth deciles exhibit a negative absolute relationship between utilisation and income, while the others have a positive one.

The results at quartile level also indicate that there is no consistent slope relationship between utilisation and income at different quartiles of income. However, unlike in the decile-level analysis, the absolute effect is always negative for the top three deciles and positive for the bottom decile.

Overall, these results differ from Lock and Sutton (2000) who found that changes in slope only occurred in the top third of the population ranked by deprivation, that is, the third most deprived.

A6.2.2 Test for a Structural Break in the Bottom Quartile

Although this test of unmet need has failed, the pattern in the quartile analysis finds that the relationship between use and income is always negative for the top three

quartiles, varying from -2.4 at the second quartile (2.75 – 5.15) to –3.76 at the third quartile and -0.74 at the top quartile, while it is positive for the bottom quartile. On that basis, we tested for a structural break between the first quartile and the rest of the sample, as reported in Table A6.2.

The variables for the lowest quartile are added to the full sample (suffixed with a '1' in Table A6.2) as well as a dummy ('poor') indicating that they are from a potentially different group from the rest of the sample. A Chow test of the joint significance of these additional variables is then applied (Greene, 1993).

Table A6.2

Test of Parameter Stability from Bottom Quartile to Other Quartiles

Variable	Coefficient	P>|t|
Age	2.90	0.00
agesq	-0.01	0.00
gender	-0.47	0.68
disabil	41.51	0.00
lonepare	5.24	0.00
marital	17.72	0.00
urbrur	0.09	0.97
numcard	-8.73	0.00
CHD	258.50	0.00
epi	266.66	0.00
rheum	307.26	0.00
diabetes	371.97	0.00
glau	262.91	0.00
respir	328.08	0.00
thyroid	103.40	0.00
psych	344.74	0.00
comor	-56.62	0.00
d2	-3.33	0.09
d3	-6.47	0.00
d4	-13.18	0.00
d5	-9.47	0.14
HB1	22.00	0.00
HB2	16.53	0.00
gpage	-7.21	0.00
gpagesq	0.07	0.00
nurse	9.69	0.01
sec	-4.26	0.39
RPA	-4.93	0.39
decpanel	0.00	0.49
Specific	182.51	0.00
symptomatic	455.60	0.00
Presum	397.57	0.00
age1	-0.17	0.58
agesq1	0.00	0.37
gender1	2.57	0.42
marital1	2.31	0.69
disabil1	14.15	0.00
lonepare1	4.43	0.26
urbrur1	-0.70	0.87

numc1	-2.04	0.07
d21	-4.29	0.43
d31	-3.30	0.50
d41	5.11	0.57
d51	-7.80	0.51
HB11	-1.66	0.69
HB21	-3.50	0.43
CHD1	11.97	0.12
epi1	34.04	0.06
rheum1	34.46	0.08
diab1	61.90	0.01
glau1	53.94	0.06
resp1	2.18	0.85
thy1	26.97	0.12
psych1	44.99	0.00
gpage1	4.16	0.11
gpagesq1	-0.05	0.07
nurse1	-6.47	0.11
sec1	0.29	0.96
RPA1	-2.02	0.69
decpanel1	0.00	0.85
spec1	-265.60	0.00
sympt1	-216.36	0.01
presum1	-378.69	0.00
comor1	-38.42	0.00
poor	157.32	0.10
_cons	-64.29	0.41
Chow Test:		0.00
R^2 (%)		38.94
N		305,887

The Chow test indicates that there is a structural break between bottom quartile and the rest of the sample. Disability, diabetes, psychiatric illness, comorbidity and the three prescribing style indicators are the variables where the bottom quartile is statistically significantly different from the rest of the sample. Note, however that given the results of Table A6.1, it is likely that a test for structural break would be accepted for other quartiles as well.

A6.2.3 Effect of An Unmet Need Adjustment

Table A6.3 illustrates the effects of measuring needs based on the top three quartiles only and applying the estimates to the full sample, versus needs estimated using the full sample. This is one of two unmet need adjustments that have been applied (Rice and Smith, 1999).

Table A6.3

Comparison of Performance Measures using Unmet Need Adjustment and Full Sample

Statistic	Unmet Need Adjustment		Full Sample	
Predicted Mean	185	(0)	185	(0)
Predicted Std. Dev	207	(0)	207	(0)
Predicted Min	-133	(1)	-136	(0)
Predicted Max	1751	(3)	1755	(2)
RMSE	268	(1)	268	(1)
MAPE	154	(0)	154	(0)
Adj – R^2	0.3877	(0.0003)	0.3893	(0.0002)
Mincer-Zarnowitz				
Const	0.866	(1.146)	-0.714	(0.794)
% of times sig.	0		0	
Slope	1.002	(0.009)	1.002	(0.005)
% of times sig.	12		20	
Concentration Index	-0.1560	(0.0001)	-0.1576	(0.0002)
Policy-Relevant Variables				
Disability	271	(0)	281	(0)
Lone Parent	66	(0)	67	(0)
Living Alone	319	(0)	320	(0)
Chronically Ill	290	(1)	296	(1)
Hardship	257	(0)	259	(0)
Asylum Seekers	76	(0)	78	(0)
Unemployment Assistance	112	(0)	114	(0)
SWA	102	(0)	105	(0)
Early School Leavers	72	(0)	74	(1)
GP-level Adj R^2	0.9027	(0.0004)	0.9041	(0.0004)
GP-level Mincer-Zarnowitz				
Const	-716	(680)	-1078	(438)
No of times sig.	0		0	
Slope	1.077	(0.013)	1.074	(0.008)
No of times sig.	100		100	
GP-level Concentration Index	-0.0629	(0.0003)	-0.0633	(0.0003)

In terms of RMSE and MAPE, there is no statistical difference between the two needs formulae, while there is a tiny difference in adjusted R^2. Equally, they are statistically indistinguishable with respect to predicted mean and standard deviation. Prediction bias is a little greater in the full sample, where the slope of a regression of actual expenditure on predicted expenditure differs from unity on 20% of occasions, as opposed to 12% following an unmet need adjustment. However, in neither case does the intercept differ from zero.

With respect to distribution of predictions, the unmet need adjustment actually has a less pro-poor concentration index than the full sample. Their predictions for vulnerable groups are very similar. At GP level, the two approaches are indistinguishable for all performance statistics except adjusted R^2, where the full sample is slightly better.

A6.3 CONCLUSION

Figure A6.1 presented prima facie evidence that there was income-related unmet need in the GMS population. However, when we tested this formally using spline regression, no such phenomenon was in evidence, with the relationship between income and utilisation differing at almost all levels of income. We did find that in the bottom quartile, not only was the slope different to the other quartiles, but the sign of the slope was different as well. Therefore, we tested this further. A Chow test of parameter stability found that the bottom quartile was different structurally from the other quartiles. However, when we compared the standard need estimates generated by the Supply model with need estimates generated following an unmet needs adjustment, we found little difference between the two. They differed marginally with respect to prediction bias and predicted mean, while they are indistinguishable with respect to all other statistics. As such, the application of an unmet need adjustment, especially given the flimsy evidence of its existence is not warranted. Future research may examine other ways of statistically testing for unmet need. For instance, if the poor are structurally different from the rest of the sample, then a two-component finite mixture model may be able to detect this structural difference.

Bibliography

J Abreveya. The effects of demographics and maternal behavior on the distribution of birth outcomes. *Empirical Economics* 2001; **26:** 247-259

JP Acton. Nonmonetary factors in the demand for medical services: some empirical evidence. *Journal of Political Economy,* 83(3):595-614, 1975.

PD Allison. Multiple Imputation for missing data: a cautionary tale. *Sociological Methods and Research*, 28:301-309, 2000.

PA Andersson, E Varde, F Diderichsen. Modelling of resource allocation to health care authorities in Stockholm county. *Health Care Management Science*, 3(2):141-149, 2000.

AS Ash, F Porell, L Gruenberg et al. Adjusting Medicare capitation payments using prior hospitalisation data. *Health Care Financing Review*, 10(4):17-29, 1989.

AE Beaton, JW Tukey. The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics*, 16:146-185, 1974.

K Beck, P Zweifel. Cream-skimming in deregulated social health insurance: evidence from Switzerland. In: P Zweifel (ed). *Health, the medical profession and regulation*, Kluwer: Dordecht, the Netherlands, 1998.

DK Blough, CW Madden, MC Hornbrook. Modelling risk using generalized linear models. *Journal of Health Economics*, 18:153 – 171, 1999.

F Breyer. *Identifying determinants of health care utilisation in a large panel data set of German sickness fund members – with application to risk-adjustment.* Paper presented at International Health Economics Association conference, University of York, July 2001.

D Brownstone, R Valletta. The bootstrap and multiple imputations: harnessing increased computing power for improved statistical tests, *Journal of Economic Perspectives*, 15(4):129-141, 2001.

D Brownstone, R Valletta. Modelling earnings measurement error: a multiple imputations approach. *Review of Economics and Statistics*, 78(4):705-17, 1996.

M Buchinsky. Changes in the US wage structure 1963-1987: application of quantile regression. *Econometrica*, 62(2):405-458, 1994.

T Callan, B Nolan, CT Whelan. *A Review of the commission on social welfare's minimum adequate income.* Policy research series paper no. 29, ESRI Dublin, December 1996.

AC Cameron, PK Trivedi, F Milne, J Piggott. A microeconometric model of demand for health care and health insurance in Australia. *Review of Economic Studies*, LV:85-106, 1988.

RA Carr-Hill, G Hardman, S Martin, S Peacock, TA Sheldon, P Smith. *A formula for distributing NHS revenues based on small area use of hospital beds.* Centre for Health Economics, University of York, York, 1994.

Central Statistics Office. *Health Statistics.* Stationery Office: Dublin, 1999.

Central Statistics Office. *Household Budget Survey 1995.* Stationery Office: Dublin, 1995.

Central Statistics Office. *Household Budget Survey 1999/2000.* Stationery Office: Dublin, 1995.

Central Statistics Office. *Census of Population, 1996.* Stationery Office: Dublin, 1998.

N Chaturvedi, Y Ben-Shlomo. From the surgery to the surgeon: does deprivation influence consultation and operations rates? *British Journal of General Practice*, 45:127-131, 1995.

CF Citro, RT Michael. *Measuring Poverty: a new approach.* National Academy Press: Washington DC, 1995.

DO Clarke, M Von Korff, K Saunders, WM Baluch, GE Simon. A chronic disease score with empirically derived weights. *Medical Care,* 33:783-95, 1995.

TJ Cole. Derivation of a needs based capitation formula for allocating prescribing budgets to health authorities and primary care groups in England: Regression analysis. Commentary: The emphasis on transparency weakens the formula. *British Medical Journal*, 320:284-288, 2000.

FA Connell, LA Bilde, MA Hanken. Clinical correlates of small area variations in population-based admission rates for diabetes. *Medical Care*, 22(10):939-949, 1984.

RD Cook. Detection of influential observations in linear regression. *Technometrics*, 19:15-18, 1977.

J Cotter. *Downside risk for European equity markets.* Paper presented at Department of Economics Visiting Speakers seminar, University College Cork, 2002.

NE Coulson, BC Stuart. Insurance choice and the demand for prescription drugs. *Southern Economic Journal*, 55:1146-1157, 1995.

FA Coutler, FA Cowell, SP Jenkins. Equivalence scale relativities and the extent of inequality and poverty. *Economic Journal*, 102:1067-1082, 1992.

FA Cowell, M Mercader-Prats. Equivalence scales and inequality. In J Silber (ed.). *Handbook on Income Inequality Measurement.* Kluwer: London, 1999.

D Cutler, L Sheiner. The geography of Medicare. *American Economic Review*, 89(2):228-233, 1999.

D Cutler, L Katz. Rising inequality? Changes in the distribution of income and consumption in the 1980s. *American Economic Review*, Papers and proceedings 82:546-551, 1992.

A Deaton. *The analysis of household surveys: a microeconometric approach to development policy*. John Hopkins University Press: Baltimore, 1997.

P Deb, PK Trivedi. Demand for medical care by the elderly: a finite mixture approach. *Journal of Applied Econometrics*, 12:313-336, 1997.

P Deb, AM Holmes, Estimates of Use and Costs of Behavioral Health Care: A Comparison of Standard and Finite Mixture Models, *Health Economics*, 9:475-489, 2000

P Deb, PK Trivedi, The Structure of Demand for Health Care: Latent Class versus Two-part Models, *Journal of Health Economics*, 21:601-625, 2002

P Deb, J Burgess. *A quasi-experimental comparison of econometric models for individual health care expenditures*, manuscript, Hunter College, City University of New York, 2002.

Department of Health. *Review of the Resource Allocation Working Party formula,* Final Report, HMSO: London, 1988.

Department of Health. *Independent enquiry into inequalities in health, Report, chaired by Sir Donald Acheson.* HMSO: London, 1999.

Department of Health and Children. *Shaping a Healthier Future*. Stationery Office: Dublin, 1993.

Department of Health and Children. *Prescribing practices and the development of general practitioner services*. Comptroller and Auditor General report on value for money examination, Stationery Office: Dublin, 1997.

Department of Health and Children. *The National Health and Lifestyle Surveys: Results from SLAN [Survey of Lifestyle, Attitudes and Nutrition] and HBSC [Health Behaviour in School-Aged Children]*. Health Promotion Unit, Department of Health and Children: Dublin, 1999.

Department of Health and Children. *Quality and Fairness. A health system for you.* Stationery Office: Dublin, 2001.

Department of Health and Social Security. *Sharing resources for health in England – Report of the Resource Allocation Working Party*, HMSO: London, 1976.

Department of Health and Social Security. *Inequalities in health. Report of a Research Working Group, chaired by Sir Douglas Black*, HMSO: London, 1980.

Department of Social Welfare. *Guide to Social Welfare Services*. Department of Social Welfare: Dublin, 1994.

Department of Social, Community and Family Affairs. *Guide to Social Welfare Services*. Department of Social, Community and Family Affairs: Dublin, 1999.

F Didrechsen, E Varde, M Whitehead. Resource allocation to health authorities: the quest for an equitable formula in Britain and Sweden. *British Medical Journal*, 315:875-8, Oct 1997.

P Diehr, D Yanez, A Ash, M Hornbrook, DY Lin. Methods for analyzing health care utilization and costs. *Annual Review of Public Health*, 20:125-144, 1999.

N Duan. Smearing estimate: a nonparametric retransformation method. *Journal of the American Statistical Association*, 78:605-610, 1983.

N Duan, WG Manning, CN Morris, JP Newhouse. A comparison of alternative models for the demand for medical care. *Journal of Business and Economic Statistics*, 1:115-126, 1983.

DL Dunn. Applications of health risk-adjustment: what can be learned from experience to date? *Inquiry,* 35: 132 – 147, 1998.

RP Ellis. *Using fuzzy diagnosis for risk-adjustment.* http://econ.bu.edu/ellis/Papers/EllisFuzzyDiag5.pdf [accessed September 2002].

B Erens P Primatesta. *Health Survey for England 1998.* Stationery Office: London, 1999.

P Fishman, DK Shay. Development and estimation of a pediatric chronic disease score using automated pharmacy data. *Medical Care*, 37(9):874-883, 1999.

S Foster, M Stano. Small area variations: a critical review of propositions, methods and evidence. *Medical Care Review*, 47:419-465, 1990.

VR Fuchs, MJ Kramer. *Determinants of expenditures for physicians' services in the United States 1948-1968.* National Bureau of Economic Research: New York, 1972.

M Gaynor, SW Polachek. Measuring information in the market: an application to physician services. *Southern Economic Journal*, 55:815-831, 1995.

P Geil, A Million, R Rotte, KF Zimmermann. Economic incentives and hospitalization in Germany. *Journal of Applied Econometrics,* 12:295-312, 1997.

F Goldman, M Grossman. The demand for health care: an hedonic approach. *Journal of Political Economy*, 86,2(1):259-281, 1978.

C Goodall. M-estimators of location: an outline of the theory. In DC Hoaglin, F Mosteller, and JW Tukey (eds.). *Understanding Robust and Exploratory Data Analysis.* Wiley: New York, 1983.

H Gravelle, M Sutton. Trends in geographical inequalities in provision of General Practitioners in England and Wales. *Lancet*, 352:1910, 1998.

H Gravelle, M Dusheiko, M Sutton. The demand for elective surgery in a public system: time and money prices in the UK National Health Service, *Journal Of Health Economics*, 21(3):423-449, 2002.

WH Greene. *Econometric Analysis*. Second Edition. Macmillan: New York, 1993.

D Grembowski, RM Andersen, M Chen. A public health model of the dental care process. *Medical Care Review* 46(4):439 - 496, 1989.

PV Grootendorst. A comparison of alternative models of prescription drug utilisation. *Health Economics,* 4:183-198, 1995.

M Grossman. *The demand for health: a theoretical and empirical investigation.* Columbia University Press: New York, 1972.

M Grossman. The human capital model. In JP Newhouse, AJ Culyer (eds). *The Handbook of Health Economics.* North Holland: Amsterdam, 2000.

J Grytten, R Sorensen. Type of contract and supplier-induced demand for primary physicians in Norway. *Journal of Health Economics*, 20:379-393, 2001.

K Hancock, DR Holden, JK Swales. Consistency of the spatial allocation of NHS hospital resources: an econometric analysis. *Applied Economics*, 23:1623 – 1636, 1991.

J Hartog, PD Pereira, JAC Vieira. Changing returns to education in Portugal during the 1980s and early 1990s: OLS and quantile regression estimators. *Applied Economics,* 33:1021-1037, 2001.

J Hausman. Mismeasured variables in econometric analysis: Problems from the right and problems from the left, *Journal of Economic Perspectives*, 15(4):57-68, 2001

R Haynes, G Bentham, A Lovett, S Gale. Effects of distances to hospital and GP surgery on hospital inpatient episodes, controlling for needs and provision. *Social Science and Medicine*, 49:425-433, 1999.

HM Hazelkorn. A comparison of dental treatment plans under different reimbursement systems. *Journal of Public Health Policy,* 6(2):223-235, 1985.

J Heckman. Sample selection bias as a specification error. *Econometrica*, 47(1):153-161, 1979.

AL Hillman. Health maintenance organisations, financial incentives and physician judgements (editorial), *Annals of Internal Medicine*, 112:891-893, 1990.

A Holly, L Gardiol, Y Eggli, T Yalcin. Diagnosis-based risk adjustment model for the Swiss health care system. Paper presented at International Health Economics Association conference, York, 2001.

J Honaker, A Joseph, G King, K Scheve, N Singh. *Amelia: A program for Missing Data.* Harvard University: Cambridge, MA, 1999. http://gking.harvard.edu [accessed May 2002].

MC Hornbrook, MJ Goodman, MD Bennett. Assessing health plan case mix in employed populations: ambulatory morbidity and prescribed drug models. In RM Scheffer, LF Rossiter (eds). *Advances in Health Economics and Health Services Research*, 12:197-232, JAI Press: Greenwich, 1991.

MC Hornbrook, MJ Goodman. Assessing relative health plan risk with the RAND-36 Health Survey. *Inquiry*, 32:56-74, 1995.

MC Hornbrook, DJ Bachman, MC O'Keefe-Rossetti, RT Meenan, PA Fishman, MJ Goodman. *Methodological Issues in Prediction Modelling for Payments Systems.* Paper presented at International Health Economics Association conference, York, 2001.

PJ Huber. Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35:73-101, 1964.

J Hunt-McCool, BF Kiker, YC Ng. Estimates of the demand for medical care under different functional forms. *Journal of Applied Econometrics*, 9: 201-218, 1994.

K Jaegher and M Jaegher. Physician behaviour with demand inducement. *Journal of Health Economics,* 19:231-258. 2000.

S Jimenez-Martin, JM Labaega, M Martinez-Granado. Latent class versus two-part models in the demand for physician services across the European Union. *Health Economics*, 11:301-321, 2002.

RE Johnson, MC Hornbrook, GA Nichols. Replicating the chronic disease score (CDS) from automated pharmacy data. *Journal of clinical epidemiology*, 47:1191-99, 1994.

A Jones. Health Econometrics. In JP Newhouse, AJ Culyer (eds). *The Handbook of Health Economics.* North Holland: Amsterdam, 2000.

J Joyce-Cooney. *A study to identify the factors that determine GP prescribing cost categorisation as calculated under the Indicative Drug Target Scheme.* MSc thesis (unpublished), Trinity College Dublin: Dublin, 1999.

N Kakwani, A Wagstaff, E van Doorslaer. Socioeconomic inequalities in health: Measurement, computation, and statistical inference. *Journal of Econometrics*, 77:87-103, 1997.

D Kenkel. Consumer health information and the demand for medical care. *The Review of Economics and Statistics*, 72, 4:587-595, 1990.

G King, J Honaker, A Joseph, K Scheve. Analyzing incomplete political science data: an alternative algorithm for multiple imputation. *American Political Science Review*, 95(1):49-69, 2001.

G King, M Tomz, J Wittenberg. Making the most of statistical analysis: improving interpretation and presentation. *American Journal of Political Science*, 44(2):341-355, 2000.

LR Klein. *The test of a model is its ability to predict,* manuscript, University of Pennsylvania, 1992.

R Koenker, G Bassett Jr. Regression Quantiles. *Econometrica*, 46(1):33-50, 1978.

R Koenker, G Bassett Jr. Robust tests for heteroscedasticity based on regression quantiles. *Econometrica*, 50(1):43-61, 1982.

R Koenker, KF Hallock. Quantile regression. *Journal of Economic Perspectives,* 15(4):143-156, 2001

R Kronick, Z Zhou, T Dreyfuss. Making risk-adjustment work for everyone. *Inquiry*, 32:41-55, Spring, 1995.

CM Lambert, NP Hurst, JF Forbes, A Lochhead, M MacLeod, G Nuki. Is day care equivalent to inpatient care for active rheumatoid arthritis? Randomised controlled clinical and economic evaluation, *British Medical Journal,* 316:965-969, 1998.

LM Lamers. Pharmacy Cost Groups. A risk adjuster for capitation payments based on the use of prescribed drugs. *Medical Care*, 37(8):824 830, 1999a.

LM Lamers. Risk-adjusted capitation based on the diagnostic cost group model: an empirical evaluation with health survey information. *Health Services Research*, 33(6): 1727-1744, 1999b.

L Lamers RJA van Vliet. *Risk adjusted capitation payments: Improving the Pharmacy Cost Group model to reduce gaming possibilities.* Paper presented at the International Health Economics Association conference, York, 2001.

DCEF Lloyd, CM Harris, DJ Roberts. Specific therapeutic group age-sex related prescribing units (STAR-PUs): weightings for analysing general practices' prescribing in England. *British Medical Journal*, 311:911-994, 1995.

RJA Little. Regression with missing X's: a review. *Journal of the American Statistical Association*, 87(420):1227-1237, 1992.

P Lock, M Sutton. *Do differences in practice prescribing costs reflect relative needs?* Paper presented at the Health Economists' Study Group meeting, Nottingham, 2000.

CA Ma, TG McGuire. Optimal health insurance and provider payment. *American Economic Review*, 87(4):685-704, 1997.

CW Madden, BP Mackay, SM Skillman, M Ciol, PK Diehr. Risk adjusting capitation: applications in employed and disabled populations. *Health Care Management Science* 3:101-109, 2000.

WG Manning, JP Newhouse, N Duan, EB Keeler, A Leibowitz, MS Marquis. Health insurance and the demand for medical care: evidence from a randomized experiment. *American Economic Review*, 77(3):251-277, 1987.

WG Manning, L Blumberg, L Moulton. The demand for alcohol: The differential response to price. *Journal of Health Economics,* 14:123-148, 1995

WG Manning. The logged dependent variable, heteroscedasticity, and the retransformation problem. *Journal of Health Economics*, 17:283-295, 1998.

S Martin, N Rice, P Smith. *Risk and the GP budget holder*. Discussion paper, No 153, Centre for Health Economics, University of York: York, 1997.

T McCarthy, Keith Davies, John Gaisford, Ullrich Hoffmeyer. *Risk-adjustment and its implications for efficiency and equity in health care systems*. Pharmaceutical Partners for Better Health Care: Basle, 1995.

MW McCracken, KD West. *Inference about predictive ability.* University of Wisconsin working paper 2001-14, University of Wisconsin, 2001.

Hugh McGavock. Some patterns of prescribing by urban general practitioners. *British Medical Journal*, 145(88):900-902, 1988.

TG McGuire. Physician Agency. In JP Newhouse, AJ Culyer (eds). *The Handbook of Health Economics.* North Holland: Amsterdam, 2000.

GJ McLachlan, T Krishnan. *The EM algorithm and extensions.* Wiley: New York, 1997.

G McLaren, M Bain. *Deprivation and health in Scotland. Insights from NHS data.* Information and Statistics Division, Scottish Executive Health Department, Edinburgh, 1998.

D Mechanic. The functions and limitations of trust in the provision of medical care. *Journal of Health Politics, Policy and Law*, 23(4):661-686, 1998.

J Mincer, V Zarnowitz. The evaluation of economic forecasts. In J Mincer (ed) *Economic Forecasts and Expectations*, National Bureau of Economic Research: New York, 1969.

J Mincer. *Schooling, Experience and Earnings*. Columbia University Press: New York, 1974

TJ Morton-Jones, MAL Pringle. Prescribing costs in dispensing practices. *British Medical Journal,* 306:1244-6, 1993.

J Mullahy. Much ado about two: reconsidering retransformation and the two-part model in health econometrics, *Journal of Health Economics*, 17:247-281, 1998.

CD Mueller. *Medicaid and the use of dental services by children.* Report 84-224, National Center for Health Services Research and Health Care Technology Assessment: Rockville MD, 1984.

KM Murphy, F Welch. Empirical age-earnings profiles. *Journal of Labour Economics*, 8(2):202-229, 1990.

MB Murphy. *A review of the indicative drugs target saving scheme: report to the Minister for Health and Children.* Department of Health and Children: Dublin, 1997.

CJL Murray, AD Lopez (eds). *The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries and risk factors in 1990 and projected to 2020.* Harvard University Press, Cambridge, Massachusetts, 1996.

JP Newhouse. Risk-adjustment: Where are we now? *Inquiry*, 35:122-131, 1998.

JP Newhouse, WP Manning, EB Keeler, EM Sloss. Adjusting capitation rates using objective health measures and prior utilisation. *Health Care Financing Review,* 10:41-54, 1989.

JP Newhouse. Reimbursing health plans and health providers: efficiency in production versus selection. *Journal of Economic Literature*, 24:1236-1263, 1996.

N Payne, C Saul. Variations in the use of cardiology services in a health authority: comparison of coronary artery revascularisation rates with prevalence of angina and coronary mortality. *British Medical Journal*, 314:257-61, 1997.

CE Phelps. Information diffusion and best practice adoption. In JP Newhouse, AJ Culyer (eds). *The Handbook of Health Economics.* North Holland: Amsterdam, 2000.

CE Phelps, C Mooney. Correction and update on 'priority setting in medical technology assessment'. *Medical Care*, 30(8):744-751, 1992.

CE Phelps, C Mooney. Variations in medical practice use: causes and consequences. In RJ Arnould, RF Rich and W White (eds). *Competitive Approaches to Health Care Reform*. The Urban Institute Press: Washington, 1993.

W Pohlmeier, V Ulrich. An econometric model of the two-part decision-making process in the demand for health care. *Journal of Human Resources*, 30:153-171, 1994.

C Propper. The demand for private health care in the UK. *Journal of Health Economics*, 19:855-876, 2000.

FD Reid, DG Cook, A Majeed. Explaining variation in hospital admission rates between general practices: cross sectional study. *British Medical Journal,* 319:98-103, 1999.

N Rice, P Smith. *Approaches to capitation and risk-adjustment in health care: An international survey.* Centre for Health Economics Discussion Paper, University of York, York, October 1999.

N Rice. *Report of study to devise a formula to assist in allocating resources for general practice prescribing expenditure within Northern Ireland.* Department of Health and Social Services, Northern Ireland Office, Belfast, 1999.

N Rice, P Dixon, DCEF Lloyd, D Roberts. Derivation of a needs based capitation formula for allocating prescribing budgets to health authorities and primary care groups in England: regression analysis, *British Medical Journal*, 320:284-8, 2000.

J Richardson. The inducement hypothesis: that doctors generate demand for their own services. In: J van der Gaag, M Perlman (eds). *Health, Economics and Health Economics*, North Holland: Oxford, 1981.

JC Robinson, HS Luft, LB Gardner, EM Morrison. A method for risk adjusting employer contributions to competing health insurance plans, *Inquiry,* 28(2):107-116, 1991.

NP Roos, LL Roos. High and low surgical rates: risk factors for area residents. *American Journal of Public Health*, 71(6):591-600, 1981.

NP Roos, LL Roos, PD Henteleff. Elective surgical rates – do high rates mean lower standards? *New England Journal of Medicine* 299(7):360-365, 1977.

MJ Rostek. *How do income distributions change in Europe?* Luxembourg Income Study working paper 240, Maxwell Citizenship School of Public Affairs, Syracuse University: Syracuse, 2000.

PJ Rousseeuw, AM Leroy. *Robust Regression and Outlier Detection.* Wiley: New York, 1987

DB Rubin. Multiple Imputation after 18+ years. *Journal of the American Statistical Association*, 91(434):473-489, 1996.

J Ryan, M Sutton, D Torgerson, M Urwin. *The effects of non-monetary costs on demand for general practice*. Paper presented to Health Economists' Study Group, Birmingham, 1999.

JMC Santos-Silva, FAG Windmeijer. Two-part multi-spell models for health care demand. IFS WP W99/2, 1999.

JL Schafer. *Analysis of Incomplete Multivariate Data.* Chapman and Hall: London, 1997.

E Schokkaert, C van de Voorde. *Risk selection and the specification of the conventional risk-adjustment formula.* Centre for Economic Studies discussion paper series, DPS 00.11, KULeuven, 2000.

A Scott and A Shiell. Do fee descriptors influence treatment choices in general practice? A multilevel discrete choice model. *Journal of Health Economics*, 16:323-342, 1997.

A Shaw, A McMunn, J Field (eds) *Scottish Health Survey 1998.* Scottish Executive Department of Health: Edinburgh, 2000.

H Sinclair. *The application of small area analysis in the planning and provision of primary and secondary care services.* Paper presented at Health Services Research conference, University College Cork, Sept 1999.

R Sorensen, J Grytten. Competition and supplier-induced demand in a health care system with fixed fees. *Health Economics*, 8:497-508, 1999.

B Starfield, J Weiner, L Mumford, D Steinwachs. Ambulatory Care Groups: A Categorization of Diagnoses For Research and Management. *Health Services Research,* 26(1):53-74, 1991.

StataCorp. *Stata Statistical Software: Release 7.0.* Stata Corporation: College Station, Texas, 2000.

A Street, AM Jones, A Furuta. Cost-sharing and pharmaceutical utilisation and expenditure in Russia. *Journal of Health Economics* 18:459-472, 1999.

M Sutton, K Major, A Walker. *An integrative approach to calculating needs based shares for GMS resources.* Paper prepared for the Steering Group and Expert Group of the Arbuthnott review of resource allocation in the NHS in Scotland, 1999a.

M Sutton, D Sutton, M Urwin. Do minority ethnic groups receive more than a fair share of General Practitioner care? Paper presented at Health Economists' Study Group conference, Birmingham, 1999b.

M Sutton, P Lock. Regional differences in health care delivery: implications for a national resource allocation formula, *Health Economics,* 9:547-559, 2000

M Tomz, J Wittenberg, G King. *CLARIFY: Software for Interpreting and Presenting Statistical Results.* Version 2.0 Harvard University: Cambridge, MA, 2001. http://gking.harvard.edu/ [accessed May 2002].

J Tudor-Hart. The inverse care law. *Lancet,* 1:405-12, 1971.

EM van Barneveld, RCJA van Vliet, WPMM van de Ven. Risk-adjusted capitation payments for catastrophic risks based on multi-year prior cots. *Health Policy,* 39:123-135, 1997.

EM van Barneveld, LM Lamers, RCJA van Vliet, WPMM van de Ven. Mandatory pooling as a supplement to risk-adjusted capitation payments in a competitive health insurance market. *Social Science and Medicine,* 47(2):223-232, 1998.

EM van Barneveld, LM Lamers, RCJA van Vleit, WPMM van de Ven. Risk sharing as a supplement to imperfect capitation: a tradeoff between selection and efficiency. *Journal of Health Economics,* 20:147-168, 2001.

WPMM van de Ven, RP Ellis. Risk-adjustment in competitive health plan markets. In JP Newhouse, AJ Culyer (eds). *The Handbook of Health Economics.* North Holland: Amsterdam, 2000.

E van Doorslaer, X Koolman, F Puffer. *Equity in the use of physician visits in OECD countries.* Presented at Measuring Up: Improving Health Systems Performance in OECD Countries conference, Ottawa, 2001.

RCJA Van Vliet, LM Lamers. The high costs of death: should plans get higher payments when members die? *Medical Care*, 36:1451-60, 1998.

M Verbeek. *A Guide to Modern Econometrics.* Wiley: Chichester, 2000.

JK Vermunt, J Magidson. *Latent Gold User's Guide*, Statistical Innovations Inc.: Belmont, Massachusetts, 2000.

M Von Korff, EH Wagner, K Saunders. A chronic disease score from automated pharmacy data. *Journal of Clinical Epidemiology*, 45:197-203, 1992.

A Wagstaff, P Paci, E van Doorslaer. On the measurement of inequalities in health. *Social Science and Medicine*, 33(3)545-557, 1991.

A Wagstaff, E van Doorslaer. Progressivity, horizontal equity and reranking in health care finance: a decomposition analysis for the Netherlands. *Journal of Health Economics*, 16:499-516, 1997.

D Watson. *Factors influencing General Practitioner attachment and implications for health service usage.* Economic and Social Research Institute seminar paper, Economic and Social Research Institute: Dublin, 1996.

JE Wennberg, A Gittelsohn. Variations in medical care among small areas. *Scientific American* 246(4):120-134, 1982.

JE Wennberg, L Blowers, R Parker, AM Gittelsohn. Changes in tonsillectomy rates associated with feedback and review. *Pediatrics*, 59(6):821-826, 1977.

GR Wilensky, LF Rossiter. The magnitude and determinants of physician-initiated visits in the United States. In: J van der Gaag, M Perlman (eds). *Health, Economics and Health Economics*. North Holland: Oxford, 1981.

A Williams. Reassessing the Global Burden of Disease. *Health Economics*, 7(1):1-8, 1998.

RPH Wilson, J Hatcher, S Barton, T Walley. Influences of practice characteristics on prescribing in fundholding and non-fundholding general practices: an observational study. *British Medical Journal*, 313:595-9, 1996.

N Woods. *Utilisation and Need in the Dental Treatment Services Scheme in Ireland.* Paper presented to PhD workshop (unpublished), Department of Economics, University College Cork: Cork, 2002.

DK Whynes, DL Baines, KH Tolley. GP fundholding and the costs of prescribing. *Journal of Public Health Medicine*, 17(3):323-329, 1995.

DK Whynes, DL Baines, KH Tolley. Prescribing costs in UK general practice: the impact of hard budget constraints. *Applied Economics* 29:393-399, 1997.

A Zaman, PJ Rousseeuw, M Orhan. Econometric application of high-breakdown robust regression techniques. *Economic Letters* ,71:1-8, 2001.

M Zimmerman-Murphy. The importance of sample selection in the estimation of medical demand equations. *Eastern Economic Journal*, 13:19-29, 1987.

P Zweifel, F Breyer. *Health Economics*. Oxford University Press: Oxford, 1997.